

XX The present sequence is human Bcl-X-like protein.
 CC The Bcl-X-like polynucleotides are useful in therapeutic, diagnostic
 CC and pharmacogenetic applications. They are useful for screening drugs
 CC effective in the treatment of symptomatic or phenotypic manifestations
 CC perturbing the normal function of protein in the body and also for
 CC treating physiological disorders and diseases. The Bcl-X-like
 CC polynucleotides are useful in conjunction with polymerase chain
 CC reaction to screen libraries, isolate clones, to prepare cloning
 CC and sequencing templates and as hybridisation probes for assessing
 CC gene expression patterns.
 XX

Sequence 327 AA:

Query Match 100.0%; Score 1709; DB 22; Length 327;
 Best Local Similarity 100.0%; Pred. No. 8.1e-166;
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCSTSGCDLEIPDDDDLTIEFKIAYTRHHVFKSTPALFSPKLLRTSRISQRLGN 60
 Db 1 mcsstsgcdleipdddddltiefkilaaylrhvfksktpalfspkllrtsrissqrlgn 60
 QY 61 CSANESWTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVYKEDSOSTPAKVSAGQRTL 120
 Db 61 csaneswtevswpornsgsekaingkksswkaffgvyeke dsostpakvsagqrtl 120
 QY 121 EYDSSHQSQMSRCLSNVEOCLHEAVDPKYSTANRVAELIVSWPPQATQAGGFSKEI 180
 Db 121 eydshsqmsrcslsnveocleheavdpkystanrvaelivswppqatqaggfkskei 180
 QY 181 FYTEGSLFQLOGHVPVASSSKKDEBQILAKIVELIKYSGDLEERKLDKRLMHFDG 240
 Db 181 fteglstfqlqghvpaasskkdeeqllakivelikysgdlerkldkrlmghfdg 240
 QY 241 LSYSVFKTTTDOVLGVDPGRGESEVKAQGFKAALVIDTAKITAIIDNHPNRRVLGFGTKY 300
 Db 241 lsysvfktttdovlmgvdpgrgeevkaagfkaalvidtakitaidnhpnmrvlgfgtky 300
 QY 301 LKENSPWIOOHGGEKILGISHEEVD 327
 Db 301 lkenfswiqghgwekllgisheevd 327

RESULT 2

AA85166 2
 ID AAB85166 standard; Protein; 327 AA.
 XX
 AC AAB85166;

QY 307-SEP-2001 (first entry)
 Db Human Bcl-G1 polypeptide.

XX Bcl-G; cancer; cancer therapy; oncogene; apoptosis; Bcl-G1; cytostatic;
 KW antiapoptotic; chromosome 12p12.3; human.
 XX
 OS Homo sapiens.

PN WO200144282-A2.

PD 21-JUN-2001.

PF 13-DEC-2000; 2000WO-US33793.

PR 14-DEC-1999; 99US-0461641.

PA (BURN-) BURHAM INST.

PI Reed JC, Godzik A;

DR WPI; 2001-398125/42.

DR N-PSDB; AA822582.

XX Novel polynucleotide encoding Bcl-G polypeptide, useful for modulating
 PT apoptosis, and for diagnosing and treating cancer -
 PS
 XX Claim 14; Fig 2; 117pp; English.

XX The invention relates to Bcl-G polypeptides and nucleic acids encoding
 CC them. The Bcl-G polypeptides can be expressed by standard recombinant
 CC methodology. Bcl-G oligonucleotides (or its anti-sense strand) and Bcl-G
 CC specific antibodies are useful for diagnosing cancer, monitoring cancer
 CC therapy or assessing prognosis of patients with cancer. The Bcl-G
 CC polypeptides are useful for modulating the activity of an oncogenic
 CC level of apoptosis mediated by the Bcl-G polypeptide. A therapeutic
 CC composition comprising the Bcl-G polypeptide, polynucleotide or antibody
 CC is useful for treating a pathology characterized by abnormal cell
 CC proliferation especially cancer. The present sequence represents a
 XX human Bcl-G1 polypeptide.

Sequence 327 AA:

Query Match 100.0%; Score 1709; DB 22; Length 327;
 Best Local Similarity 100.0%; Pred. No. 8.1e-166;
 Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCSTSGCDLEIPDDDDLTIEFKIAYTRHHVFKSTPALFSPKLLRTSRISQRLGN 60
 Db 1 mcsstsgcdleipdddddltiefkilaaylrhvfksktpalfspkllrtsrissqrlgn 60
 QY 61 CSANESWTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVYKEDSOSTPAKVSAGQRTL 120
 Db 61 csaneswtevswpornsgsekaingkksswkaffgvyeke dsostpakvsagqrtl 120
 QY 121 EYDSSHQSQMSRCLSNVEOCLHEAVDPKYSTANRVAELIVSWPPQATQAGGFSKEI 180
 Db 121 eydshsqmsrcslsnveocleheavdpkystanrvaelivswppqatqaggfkskei 180
 QY 181 FYTEGSLFQLOGHVPVASSSKKDEBQILAKIVELIKYSGDLEERKLDKRLMHFDG 240
 Db 181 fteglstfqlqghvpaasskkdeeqllakivelikysgdlerkldkrlmghfdg 240
 QY 241 LSYSVFKTTTDOVLGVDPGRGESEVKAQGFKAALVIDTAKITAIIDNHPNRRVLGFGTKY 300
 Db 241 lsysvfktttdovlmgvdpgrgeevkaagfkaalvidtakitaidnhpnmrvlgfgtky 300
 QY 301 LKENSPWIOOHGGEKILGISHEEVD 327
 Db 301 lkenfswiqghgwekllgisheevd 327

RESULT 3

AAE07041 3
 ID AAE07041 standard; Protein; 252 AA.
 XX
 AC AAE07041;

QY 23-OCT-2001 (first entry)
 Db Human BCL-X-like protein #2.

XX Human; BCL-X-like protein; therapy; physiological disorder.
 KW
 XX
 OS Homo sapiens.

PN WO200157213-A2.

PD 09-AUG-2001.

PF 31-JAN-2001; 2001WO-US03446.

PR 04-FEB-2000; 2000US-0180412.

Thu Jun 20 13:14:55 2002

us-09-771-961-2.rag

PA (LEXI-) LEXICON GENETICS INC.
 XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A, Zambrowicz B;
 PI Sands AT;
 XX WPI: 2001-488882/53.
 DR N-PSDB: AAD13236.
 XX Novel isolated human BCL-X-like polynucleotide, useful in therapeutic,
 PT diagnostic and pharmacogenic applications
 PS Claim 5; Page 32; 33pp; English.
 CC The present sequence is human BCL-X-like protein.
 CC The BCL-X-like polynucleotides are useful in therapeutic, diagnostic
 CC and pharmacogenic applications. They are useful for screening drugs
 CC effective in the treatment of symptomatic or phenotypic manifestations
 CC perturbing the normal function of protein in the body and also for
 CC treating physiological disorders and diseases. The BCL-X-like
 CC polynucleotides are useful in conjunction with polymerase chain
 CC reaction to screen libraries, isolate clones, to prepare cloning
 CC and sequencing templates and as hybridisation probes for assessing
 CC gene expression patterns.
 CC
 SQ Sequence 252 AA:

Query Match 69.6%; Score 1189.5; DB 22; Length 252;
 Best Local Similarity 93.5%; Pred. No. 6.4e-113;
 Matches 232; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 1 MCSTSGCDLEFIPDDDDNTIEFKILAYTRHNVKSPALFSPKILRTSLSGRGLGN 60
 1 mcsstsgcdleipdddddntiefkilaaytrhvkspalfspkllrtslsgrglgn 60
 DB 61 CSANESWTEVSWPCRNSSSEKAINLGKKSSWKAFFGVVEKEDSOSTPAKYSAQGRFL 120
 61 csaneswtevwpcrnsssekaingksswkaaffgvvekedsgstpkysaagqrfl 120
 QY 121 EYQDSHSQOOSRCLSNVBOCLEHEAVDPKVIYINRYVAEIVYSWPPQATQAGFKSKEL 180
 121 eyqdshtsqosrclsnvboacleheavdpkvisianryvaeiyswppqatqagfkskel 180
 DB 121 eyqdshtsqosrclsnvboacleheavdpkvisianryvaeiyswppqatqagfkskel 180
 121 eyqdshtsqosrclsnvboacleheavdpkvisianryvaeiyswppqatqagfkskel 180
 QY 181 FVTEGLSFOLQGHVPYASSSKKDEEQLIAKIVELLKYSQGLERKRLKK-----DKAL 233
 181 fvtelgsfolqghvpyassskkdeeqliakivelkysqglertkrlk-----dkal 233
 DB 181 fvtelgsfolqghvpyassskkdeeqliakivelkysqglertkrlk-----dkal 233
 181 fvtelgsfolqghvpyassskkdeeqliakivelkysqglertkrlk-----dkal 233
 QY 234 MGHFODGL 241
 234 mghfodgl 241
 DB 241 qgfpgdgl 248

RESULT 4
 AAB85167 standard; Protein: 252 AA.
 ID AAB85167
 XX AAB85167;
 AC 07-SEP-2001 (first entry)
 DT 07-SEP-2001 (first entry)
 XX Human Bcl-Gs polypeptide.
 DE Human Bcl-Gs polypeptide.
 XX Bcl-G; cancer; cancer therapy; oncogene; apoptosis; Bcl-Gs; cytostatic;
 KW antiapoptotic; chromosome 12p12.3; human.
 KM antiapoptotic; chromosome 12p12.3; human.
 XX Homo sapiens.
 OS Homo sapiens.
 XX MO200144282-A2.
 PN 21-JUN-2001.
 XX 13-DEC-2000; 2000WO-US33793.
 PF
 XX

PR 14-DEC-1999; 99US-0461641.
 XX (BURN-) BURHAM INST.
 PA Reed JC, Godzik A;
 XX WPI: 2001-398125/42.
 DR N-PSDB: AAH22583.
 XX Novel polynucleotide encoding Bcl-G polypeptide, useful for modulating
 PT apoptosis, and for diagnosing and treating cancer
 PS Claim 14; Fig 4; 117pp; English.
 CC The invention relates to Bcl-G polypeptides and nucleic acids encoding
 CC them. The Bcl-G polypeptides can be expressed by standard recombinant
 CC methodology. Bcl-G oligonucleotides (or its anti-sense strand) and Bcl-G
 CC specific antibodies are useful for diagnosing cancer, monitoring cancer
 CC therapy or assessing prognosis of patients with cancer. The Bcl-G
 CC polypeptides are useful for modulating the activity of an oncogenic
 CC polypeptide. They are useful for identifying modulators, a therapeutic
 CC a level of apoptosis mediated by the Bcl-G polypeptide. A therapeutic
 CC composition comprising the Bcl-G polypeptide, polynucleotide or antibody
 CC is useful for treating a pathology characterized by abnormal cell
 CC proliferation especially cancer. The present sequence represents a
 CC human Bcl-Gs polypeptide.
 CC
 SQ Sequence 252 AA:

Query Match 69.6%; Score 1189.5; DB 22; Length 252;
 Best Local Similarity 93.5%; Pred. No. 6.4e-113;
 Matches 232; Conservative 2; Mismatches 7; Indels 7; Gaps 1;

QY 1 MCSTSGCDLEFIPDDDDNTIEFKILAYTRHNVKSPALFSPKILRTSLSGRGLGN 60
 1 mcsstsgcdleipdddddntiefkilaaytrhvkspalfspkllrtslsgrglgn 60
 DB 61 CSANESWTEVSWPCRNSSSEKAINLGKKSSWKAFFGVVEKEDSOSTPAKYSAQGRFL 120
 61 csaneswtevwpcrnsssekaingksswkaaffgvvekedsgstpkysaagqrfl 120
 QY 121 EYQDSHSQOOSRCLSNVBOCLEHEAVDPKVIYINRYVAEIVYSWPPQATQAGFKSKEL 180
 121 eyqdshtsqosrclsnvboacleheavdpkvisianryvaeiyswppqatqagfkskel 180
 DB 121 eyqdshtsqosrclsnvboacleheavdpkvisianryvaeiyswppqatqagfkskel 180
 121 eyqdshtsqosrclsnvboacleheavdpkvisianryvaeiyswppqatqagfkskel 180
 QY 181 FVTEGLSFOLQGHVPYASSSKKDEEQLIAKIVELLKYSQGLERKRLKK-----DKAL 233
 181 fvtelgsfolqghvpyassskkdeeqliakivelkysqglertkrlk-----dkal 233
 DB 181 fvtelgsfolqghvpyassskkdeeqliakivelkysqglertkrlk-----dkal 233
 181 fvtelgsfolqghvpyassskkdeeqliakivelkysqglertkrlk-----dkal 233
 QY 234 MGHFODGL 241
 234 mghfodgl 241
 DB 241 qgfpgdgl 248

RESULT 5
 AAB85188 standard; Protein: 328 AA.
 ID AAB85188
 XX AAB85188;
 AC 07-SEP-2001 (first entry)
 DT 07-SEP-2001 (first entry)
 XX Mouse Bcl-G polypeptide.
 DE Mouse Bcl-G polypeptide.
 XX Bcl-G; cancer; cancer therapy; oncogene; apoptosis; cytostatic;
 KW antiapoptotic; chromosome 12p12.3; mouse.
 KM antiapoptotic; chromosome 12p12.3; mouse.
 XX Mus sp.
 OS Mus sp.
 XX Key Location/Qualifiers
 FH Misc-difference 107
 FT /label= unknown
 FT

```
/note="encoded by wca"
```

W0200144282-A2.

21-JUN-2001.

13-DEC-2000; 2000WO-US33793.

14-DEC-1999; 99US-0461641.

(BURN-) BURNHAM INST.

Reed JC, Godzik A;

NPI; 2001-398125/42.

... 22670.

claim 14; Page 114-115, 117-118, 119-120, 121-122, 123-124, 125-126, 127-128, 129-130, 131-132, 133-134, 135-136, 137-138, 139-140, 141-142, 143-144, 145-146, 147-148, 149-150, 151-152, 153-154, 155-156, 157-158, 159-160, 161-162, 163-164, 165-166, 167-168, 169-170, 171-172, 173-174, 175-176, 177-178, 179-180, 181-182, 183-184, 185-186, 187-188, 189-190, 191-192, 193-194, 195-196, 197-198, 199-200, 201-202, 203-204, 205-206, 207-208, 209-210, 211-212, 213-214, 215-216, 217-218, 219-220, 221-222, 223-224, 225-226, 227-228, 229-230, 231-232, 233-234, 235-236, 237-238, 239-240, 241-242, 243-244, 245-246, 247-248, 249-250, 251-252, 253-254, 255-256, 257-258, 259-260, 261-262, 263-264, 265-266, 267-268, 269-270, 271-272, 273-274, 275-276, 277-278, 279-280, 281-282, 283-284, 285-286, 287-288, 289-290, 291-292, 293-294, 295-296, 297-298, 299-300, 301-302, 303-304, 305-306, 307-308, 309-310, 311-312, 313-314, 315-316, 317-318, 319-320, 321-322, 323-324, 325-326, 327-328, 329-330, 331-332, 333-334, 335-336, 337-338, 339-340, 341-342, 343-344, 345-346, 347-348, 349-350, 351-352, 353-354, 355-356, 357-358, 359-360, 361-362, 363-364, 365-366, 367-368, 369-370, 371-372, 373-374, 375-376, 377-378, 379-380, 381-382, 383-384, 385-386, 387-388, 389-390, 391-392, 393-394, 395-396, 397-398, 399-400, 401-402, 403-404, 405-406, 407-408, 409-410, 411-412, 413-414, 415-416, 417-418, 419-420, 421-422, 423-424, 425-426, 427-428, 429-430, 431-432, 433-434, 435-436, 437-438, 439-440, 441-442, 443-444, 445-446, 447-448, 449-450, 451-452, 453-454, 455-456, 457-458, 459-460, 461-462, 463-464, 465-466, 467-468, 469-470, 471-472, 473-474, 475-476, 477-478, 479-480, 481-482, 483-484, 485-486, 487-488, 489-490, 491-492, 493-494, 495-496, 497-498, 499-500, 501-502, 503-504, 505-506, 507-508, 509-510, 511-512, 513-514, 515-516, 517-518, 519-520, 521-522, 523-524, 525-526, 527-528, 529-530, 531-532, 533-534, 535-536, 537-538, 539-540, 541-542, 543-544, 545-546, 547-548, 549-550, 551-552, 553-554, 555-556, 557-558, 559-560, 561-562, 563-564, 565-566, 567-568, 569-570, 571-572, 573-574, 575-576, 577-578, 579-580, 581-582, 583-584, 585-586, 587-588, 589-590, 591-592, 593-594, 595-596, 597-598, 599-600, 601-602, 603-604, 605-606, 607-608, 609-610, 611-612, 613-614, 615-616, 617-618, 619-620, 621-622, 623-624, 625-626, 627-628, 629-630, 631-632, 633-634, 635-636, 637-638, 639-640, 641-642, 643-644, 645-646, 647-648, 649-650, 651-652, 653-654, 655-656, 657-658, 659-660, 661-662, 663-664, 665-666, 667-668, 669-670, 671-672, 673-674, 675-676, 677-678, 679-680, 681-682, 683-684, 685-686, 687-688, 689-690, 691-692, 693-694, 695-696, 697-698, 699-700, 701-702, 703-704, 705-706, 707-708, 709-710, 711-712, 713-714, 715-716, 717-718, 719-720, 721-722, 723-724, 725-726, 727-728, 729-730, 731-732, 733-734, 735-736, 737-738, 739-740, 741-742, 743-744, 745-746, 747-748, 749-750, 751-752, 753-754, 755-756, 757-758, 759-760, 761-762, 763-764, 765-766, 767-768, 769-770, 771-772, 773-774, 775-776, 777-778, 779-780, 781-782, 783-784, 785-786, 787-788, 789-790, 791-792, 793-794, 795-796, 797-798, 799-800, 801-802, 803-804, 805-806, 807-808, 809-810, 811-812, 813-814, 815-816, 817-818, 819-820, 821-822, 823-824, 825-826, 827-828, 829-830, 831-832, 833-834, 835-836, 837-838, 839-840, 841-842, 843-844, 845-846, 847-848, 849-850, 851-852, 853-854, 855-856, 857-858, 859-860, 861-862, 863-864, 865-866, 867-868, 869-870, 871-872, 873-874, 875-876, 877-878, 879-880, 881-882, 883-884, 885-886, 887-888, 889-890, 891-892, 893-894, 895-896, 897-898, 899-900, 901-902, 903-904, 905-906, 907-908, 909-910, 911-912, 913-914, 915-916, 917-918, 919-920, 921-922, 923-924, 925-926, 927-928, 929-930, 931-932, 933-934, 935-936, 937-938, 939-940, 941-942, 943-944, 945-946, 947-948, 949-950, 951-952, 953-954, 955-956, 957-958, 959-960, 961-962, 963-964, 965-966, 967-968, 969-970, 971-972, 973-974, 975-976, 977-978, 979-980, 981-982, 983-984, 985-986, 987-988, 989-990, 991-992, 993-994, 995-996, 997-998, 999-1000, 1001-1002, 1003-1004, 1005-1006, 1007-1008, 1009-1010, 1011-1012, 1013-1014, 1015-1016, 1017-1018,

the invention relates to a

The Bcl-2 protein is a Bcl-2 family protein and a nuclear acid encoding protein. The Bcl-2 protein can be expressed by standard recombinant technology. Bcl-2 oligonucleotides (or its anti-sense strand) and Bcl-2 specific antibodies are useful for diagnosing cancer, monitoring cancer therapy or assessing prognosis of patients with cancer. The Bcl-2 oligonucleotides are useful for modulating the activity of an oncogenic protein, Bcl-2, and Bcl-2 protein. They are useful for identifying modulators, for modulating level of apoptosis mediated by the Bcl-2 protein. A therapeutic composition comprising the Bcl-2 oligonucleotide, a therapeutic agent useful for treating a pathology characterized by abnormal cell proliferation especially cancer. The present sequence represents a

Match	54.4%	Score 929.5	DB 22	Length 328;
Local Similarity	59.5%	Pred. No. 3.6e-86;		
200; Conservative	36;	Mismatches 87		

[illegible]

378 standard; Protein; 151 AA

378;

DT	21-NOV-2001	(first entry)
XX		
DE		Human reproductive system related antigen SEQ ID NO: 4036.
XX		
KW		Human; reproductive system related antigen; reproductive system disorder.
XX		
OS		Homo sapiens.
XX		
PN	WO200155320-A2.	
XX		
PD	02-AUG-2001.	
XX		
PF	17-JAN-2001; 2001WO-US01339.	
XX		

PR	31-JAN-2000;	2000US-0179665
PR	04-FEB-2000;	2000US-0180628
PR	24-FEB-2000;	2000US-0184664
PR	02-MAR-2000;	2000US-0186350
PR	16-MAR-2000;	2000US-0188874
PR	17-MAR-2000;	2000US-0190875
PR	18-APR-2000;	2000US-0190676
PR	19-MAY-2000;	2000US-0196123
PR	07-JUN-2000;	2000US-0205515
PR	28-JUN-2000;	2000US-0209467
PR	30-JUN-2000;	2000US-0214886
PR	07-JUL-2000;	2000US-0215135
PR	07-JUL-2000;	2000US-0216647
PR	11-JUL-2000;	2000US-0216880
PR	11-JUL-2000;	2000US-0217487
PR	14-JUL-2000;	2000US-0217496
PR	26-JUL-2000;	2000US-0218290
PR	26-JUL-2000;	2000US-0220963
PR	14-AUG-2000;	2000US-0220964
PR	14-AUG-2000;	2000US-0224518
PR	14-AUG-2000;	2000US-0224519
PR	14-AUG-2000;	2000US-0228213
PR	14-AUG-2000;	2000US-0228214
PR	14-AUG-2000;	2000US-0228266
PR	14-AUG-2000;	2000US-0228267
PR	14-AUG-2000;	2000US-0228266
PR	14-AUG-2000;	2000US-0225447
PR	14-AUG-2000;	2000US-0225447
PR	14-AUG-2000;	2000US-0225758
PR	14-AUG-2000;	2000US-0225757
PR	18-AUG-2000;	2000US-0225759
PR	22-AUG-2000;	2000US-0226681
PR	22-AUG-2000;	2000US-0226681
PR	23-AUG-2000;	2000US-0227009
PR	30-AUG-2000;	2000US-0227009
PR	01-SEP-2000;	2000US-0229624
PR	01-SEP-2000;	2000US-0229624
PR	01-SEP-2000;	2000US-0229343
PR	01-SEP-2000;	2000US-0229344
PR	03-SEP-2000;	2000US-0229345
PR	05-SEP-2000;	2000US-0229509
PR	05-SEP-2000;	2000US-0229509
PR	06-SEP-2000;	2000US-0230437
PR	08-SEP-2000;	2000US-0230438
PR	08-SEP-2000;	2000US-0231242
PR	08-SEP-2000;	2000US-0231243
PR	08-SEP-2000;	2000US-0231244
PR	08-SEP-2000;	2000US-0231413
PR	08-SEP-2000;	2000US-0231413
PR	12-SEP-2000;	2000US-0232080
PR	12-SEP-2000;	2000US-0232081
PR	14-SEP-2000;	2000US-0233967
PR	14-SEP-2000;	2000US-0233968
PR	14-SEP-2000;	2000US-0233968
PR	14-SEP-2000;	2000US-0233969
PR	14-SEP-2000;	2000US-0233401
PR	14-SEP-2000;	2000US-0233401
PR	14-SEP-2000;	2000US-0233063

us-09-771-961-2.rag

Thu Jun 20 13:14:55 2002

PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 21-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234999.
PR 26-SEP-2000; 2000US-0235834.
PR 26-SEP-2000; 2000US-0235835.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0235837.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 29-SEP-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 02-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0246176.
PR 20-OCT-2000; 2000US-0246177.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 08-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249220.
PR 17-NOV-2000; 2000US-0249221.
PR 17-NOV-2000; 2000US-0249222.
PR 17-NOV-2000; 2000US-0249223.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249225.
PR 17-NOV-2000; 2000US-0249226.
PR 17-NOV-2000; 2000US-0249227.
PR 17-NOV-2000; 2000US-0249228.
PR 17-NOV-2000; 2000US-0249229.
PR 17-NOV-2000; 2000US-0249230.
PR 17-NOV-2000; 2000US-0249231.
PR 17-NOV-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251890.
PR 08-DEC-2000; 2000US-0251990.
PR 08-DEC-2000; 2000US-0254097.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX N-PSDB; AAL01348.
XX WPI; 2001-465570/50.
DR Isolated nucleic acid molecule encoding a reproductive system antigen -
XX N-PSDB; AAL01348.
XX Is used in preventing, treating or ameliorating a medical condition.
PT Claim 11: SEQ ID NO 4036; 1297pp + Sequence Listing: English.
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a protein of the invention.
XX Sequence 151 AA;
SQ

Query Match 40.5%; Score 691.5; DB 22; Length 151;
Best Local Similarity 95.0%; Pred. No. 2, 2e-62; Indels 1; Gaps 1;
Matches 134; Conservative 1; Mismatches 5;
1 MCSTSGCLEIEIPDDDDIETFEKILAYTRHHVFKSTPAISPKILRTSRISORIGIN 60
10 mcsstsgcleieipddddlttletfklaytrhhvfkscpalisfpllrtsisqgqir 69
QY 61 CSANESWTEVSPCRNSQSEKAINIGKKSSMKAFEGVEKEQSOPAPKVSNOGORTL 129
QY 70 csanewtevwpcrnsqsekaingkksswkalfgyvekedsqspkavsaqgqir 129
Db 121 EYQDSHQMSRCLSNVQCL 141
QY 130 eyqdsqss-gqylsnveqvl 149
Db

RESULT 7
ABB27961 standard; Peptide; 129 AA.
ID ABB27961
AC ABB27961;
DT 01-FEB-2002 (first entry)
DE Human peptide #612 encoded by breast cell single exon nucleic acid probe.
DE Human peptide #612 encoded by breast cell single exon nucleic acid probe.
KW Human; microarray; single exon probe; gene expression; breast;
OS Homo sapiens.
XX disease; cancer.
XX PN WC200157271-R2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00662.
XX PP 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632266.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
WPI: 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,
XX useful for measuring gene expression in sample derived from human
XX breast, comprises number of single exon nucleic acid probes -
XX Claim 27; SEQ ID NO 10929; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting
XX the probes with a collection of detectably labeled nucleic acids
XX bound to each probe of the microarray, and then measuring the label
XX verifying the expression of regions of genomic DNA predicted to
XX encode proteins. They are useful for gene discovery, and for
XX determining predisposition and/or prognosis breast, and for
XX agents on cells. The microarray of this invention presents a far greater
XX than expressed sequence tag microarrays. The method is suitable for
XX rapid production of functional information from genomic sequence. The
XX present sequence is a peptide encoded by a single exon nucleic acid
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
SQ Sequence 129 AA;

Query Match
Best Local Similarity 40.1%; Score 686; DB 22; Length 129;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 DDLNTEKILAYTTRHVFSTPALSPKILTRISLGRGNCNSANESWTEVSMPCR 75
Db 1 dddlnlteklilaytrhvfstpalspkiltrislsgrgncnsaneswtevsmpcr 75
Oy 76 NSQSEKAINLGKKSSWKAFFGVVEKEDSOSTPAKVSAGQRTLEYDSDSHSOQSRCLS 135
Db 61 nsqsekainglkksswkaffgvvekedsgstpakvsagqgtrleydshsqswrcis 120
Oy 136 NVEQCLEHE 144
Db 121 nveqclehe 129

RESULT 8

ABB3133
ID ABB3133 standard; Peptide; 129 AA.
AC ABB3133;

04-FEB-2002 (first entry)

Peptide #639 encoded by human foetal liver single exon probe.
Human; foetal liver; gene expression; single exon nucleic acid probe.

Homo sapiens.
WO200157277-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00669.

04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.
PR 03-SEP-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
WPI: 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 25768; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
SQ Sequence 129 AA;

Query Match
Best Local Similarity 40.1%; Score 686; DB 22; Length 129;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 DDLNTEKILAYTTRHVFSTPALSPKILTRISLGRGNCNSANESWTEVSMPCR 75
Db 1 dddlnlteklilaytrhvfstpalspkiltrislsgrgncnsaneswtevsmpcr 75
Oy 76 NSQSEKAINLGKKSSWKAFFGVVEKEDSOSTPAKVSAGQRTLEYDSDSHSOQSRCLS 135
Db 61 nsqsekainglkksswkaffgvvekedsgstpakvsagqgtrleydshsqswrcis 120
Oy 136 NVEQCLEHE 144
Db 121 nveqclehe 129

RESULT 9

ABB18598
ID ABB18598 standard; Protein; 129 AA.
AC ABB18598;

23-JAN-2002 (first entry)

Protein #597 encoded by probe for measuring heart cell gene expression.
Human; gene expression; heart; microarray; vascular system;
cardiovascular disease; hypertension; cardiac arrhythmia;

congenital heart disease.
Homo sapiens.

WO200157274-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00666.

04-FEB-2000; 2000US-0180312.
26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.

	Sequence	Conservative	0;	Mismatches	0;	Indels	0;	Gaps
Oy	16 DDILNIIEKILAIYTRHNEKSPALFSEKILRTFSISORGIGMCSANESMTIEVMPER	0;						
Db	1 dddltltiekkllaytrthvfkscpalifpklrltrstlsqrgyncsaneatevswmpor	0;						
Oy	76 NQSEKAINIGKSKSKMAFGEVEKEEDSOTPAKVSAGQORTLEKODSHSQMSRCLS	0;						
Db	61 nsgsekaalnlgkkskswaalfgvvekedsgsfprakvsagqgltleyqshsqswrcsls	0;						
Oy	136 NVEGCLLEH 144	0;						
Db	121 nvegclehe 129	0;						

AA Peptide #620 (first entry)
DE
XX
XX Probe; human; microarray; gene expression
XX cervical cancer.
XX Homo sapiens.
XX cervical epithelial cell;

XX	MO200157278-72.
FN	09-AUG-2001.
ED	
XX	
XX	30-JAN-2001; 2001MO-US00670.
PR	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
XX	30-JUN-2000; 2000US-0207456.
PR	03-AUG-2000; 2000US-0620408.
PR	21-SEP-2000; 2000US-0632366.
PR	27-SEP-2000; 2000US-0234667.
XX	04-OCT-2000; 2000US-0236359.
XX	2000GB-0024263.
XX	
XX	(MOLE-) MOLECULAR DYNAMICS INC.
PI	
XX	Penn SG, Hanzel DK, Chen W, Rank DR.
XX	WPI; 2001-488901/53.

Query Match	40.1%	Score 686;	DB 22:	Length 129;	
Best Local Similarity	100.0%	Pred. No. 6, 3e-62;			
Matches 129;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	16	DDDLNTEKIIAYVTRHVFKKSPALFESKRLTPTSLSORIGCSANESWTEVSPGR			
Db	1	dddlntetkllayvtrthvfekstpalrfspkllrtlrtsqglncsaanswlevsmpcr			
QY	76	NSOSSEKAINIGKKRSWKAFFGVVLEKEDSOSTPAKSAAGCOORTLEOXDSHSGQMSRCLS			
Db	61	nsqsekaalnlgkkskwkaffgvvlekedsgostpakvsagqgrtleyqdsqshsqsrcls			
QY	136	NVEQCLLEHE 144			
Db	121	nveqclele 129			

AA	RESULT_13
AM	AAAM6596
ID	AAAM6596 standard; Protein; 129 AA.
XX	
AC	AAAM6596;
XX	
DT	17-Oct-2001 (first entry)
DE	
XX	
XX	Peptide #63 encoded by probe for measuring placental gene expression.
XX	Probe: microarray; human; placenta; antenatal diagnosis.
XX	
XX	Homo sapiens.
SS	

PN MG200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US000603.
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0207456.
PR 03-AUG-2000; 2000US-0608408.
PR 21-SEP-2000; 2000US-0633366.
PR 27-SEP-2000; 2000US-0234667.
PR 04-OCT-2000; 2000US-0236359.
PR 20-OCT-2000; 2000GB-0024263.
PA (MOLE-) MOLECULAR DYNAMICS INC.
PI Penn SG, Hanzel DK, Chen W, Rank DR
XX WPI; 2001-468897/53.
XX Human genome-3.
XX

genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -

XX Claim 27: SEQ ID No 26865; 654pp; English.
 XX The present invention relates to single exon nucleic acid probes (SENPs).
 CC see A113135-A115346). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.

XX Sequence 129 AA:

Query Match 40.1%; Score 686; DB 22; Length 129;
 Best Local Similarity 100.0%; Pred. No. 6, 3e-62; Indels 0; Gaps 0;
 Matches 129; Conservative 0; Mismatches 0

16 DDDINTIEKILAYTRHHVFKSPALFSPKILRTSLISORGLNCANESWTEVSPCR 75
 1 dddintieffilaytrhhvfkstpalispllrtslsqrglncaneswtevsqpcr 60
 QY 76 NSOSSEKAINLGRKSSWKAFEGVKEKEDSSTPAKVSNOGORTLEYODSHSQWSCRCL 135
 1 nssossekaingrksswkafegvkekedsstpakvsnogortleyodshsqwscrcls 120
 Db 61 nssossekaingrksswkafegvkekedsstpakvsnogortleyodshsqwscrcls 120
 QY 136 NVEQCLEHE 144
 1 nveqclehe 129
 Db 121 nveqclehe 129

RESULT 14

AA01922 standard; Protein: 129 AA.

AA01922:

09-OCT-2001 (first entry)

Peptide #604 encoded by probe for measuring human breast gene expression.

Probe: human; breast disease; breast cancer; development disorder;

Inflammatory disease; proliferative breast disease; non-carcinoma tumour.

Homo sapiens.

WO200157270-A2.

09-AUG-2001.

29-JAN-2001; 2001WO-US00661.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0612386.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression

in a human breast -

Claim 27: SEQ ID No 10662; 322pp; English.

The present invention relates to novel single exon nucleic acid probes

(see A110010-A110067). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases of development,
 CC aetiology. The diseases include: breast cancer, disorders of proliferative
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 129 AA:

Query Match 40.1%; Score 686; DB 22; Length 129;
 Best Local Similarity 100.0%; Pred. No. 6, 3e-62; Indels 0; Gaps 0;
 Matches 129; Conservative 0; Mismatches 0

16 DDDINTIEKILAYTRHHVFKSPALFSPKILRTSLISORGLNCANESWTEVSPCR 75
 1 dddintieffilaytrhhvfkstpalispllrtslsqrglncaneswtevsqpcr 60
 QY 76 NSOSSEKAINLGRKSSWKAFEGVKEKEDSSTPAKVSNOGORTLEYODSHSQWSCRCL 135
 1 nssossekaingrksswkafegvkekedsstpakvsnogortleyodshsqwscrcls 120
 Db 61 nssossekaingrksswkafegvkekedsstpakvsnogortleyodshsqwscrcls 120
 QY 136 NVEQCLEHE 144
 1 nveqclehe 129
 Db 121 nveqclehe 129

RESULT 15

ABG26266 standard; Protein: 178 AA.

ABG26266:

18-FEB-2002 (first entry)

Novel human diagnostic protein #26257.

Human; chromosome mapping; gene mapping; gene therapy; forensic;

Food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Dremanac RF, Liu C, Tang YT;

WPI: 2001-639362/73.

N-PSDB: AAS90453.

New isolated polynucleotide and encoded polypeptides, useful in

diagnostics, forensics, gene mapping, identification of mutations

responsible for genetic disorders or other traits and to assess

biodiversity -

Claim 20: SEQ ID No 56625; 103pp; English.

The invention relates to isolated polynucleotide (I) and

polypeptide (II) sequences. (I) is useful as hybridisation probes,

polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. Abc00010-Abc30377 represent novel human
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 178 AA;

RY Match 35.6%; Score 608; DB 22; Length 178;
 Best Local Similarity 100.0%; Pred. No. 9.6e-54;
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 211 KIVELLYSGDLEKTKKALMGHFDGLSVFKTTDOVLMGVDPGESEVKAQGF 270
 Db 2 KIVELLYSGDLEKTKKALMGHFDGLSVFKTTDOVLMGVDPGESEVKAQGF 270
 OY 271 KALVLDVYAKLAIIDNHPNRLVLCFGTKYLNKENSFWIOHGGWKEKILGTSHEVD 327
 Db 62 kaalvldvtakilaiddnphmrivlgfgykykenfswlqhgqgweklgishvevd 118

Search completed: June 19, 2002, 16:34:38
 Job time: 131 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2002, 16:34:38 ; Search time 53.19 Seconds
(without alignments)
526.238 Million cell updates/sec

Title: US-09-771-961-4
Perfect score: 1316
Sequence: 1 MCSRSGCDLEIPDDDLN.....IPVDTSTGCPQGMACI 252

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Archived: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A-Genesec_032802:*

- 1: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/hold-genesec/genesecp-emb1/AA2001.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1316	100.0	252	22	AAE07041	Human BCL-X-like p
2	1316	100.0	252	22	AAE07041	Human BCL-X-like p
3	1189.5	90.4	327	22	AAE07040	Human BCL-X-like p
4	1189.5	90.4	327	22	AAE07040	Human BCL-X-like p
5	691.5	52.5	151	22	AAE05378	Human Bcl-Gl Polyp
6	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
7	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
8	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
9	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
10	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
11	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp

12	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
13	686	52.1	129	22	AAE05378	Human Bcl-Gl Polyp
14	501	38.1	328	22	AAE05378	Human Bcl-Gl Polyp
15	93	7.1	1250	22	AAE05378	Human Bcl-Gl Polyp
16	91.5	7.0	1075	19	AAE05378	Human Bcl-Gl Polyp
17	91.5	7.0	1075	19	AAE05378	Human Bcl-Gl Polyp
18	91.5	7.0	1075	19	AAE05378	Human Bcl-Gl Polyp
19	90	6.8	660	22	AAE05378	Human Bcl-Gl Polyp
20	89.5	6.8	660	22	AAE05378	Human Bcl-Gl Polyp
21	89	6.8	660	22	AAE05378	Human Bcl-Gl Polyp
22	89	6.8	660	22	AAE05378	Human Bcl-Gl Polyp
23	88.5	6.7	1369	22	AAE05378	Human Bcl-Gl Polyp
24	88.5	6.7	1369	22	AAE05378	Human Bcl-Gl Polyp
25	88.5	6.7	1369	22	AAE05378	Human Bcl-Gl Polyp
26	86	6.5	635	22	AAE05378	Human Bcl-Gl Polyp
27	85.5	6.5	635	22	AAE05378	Human Bcl-Gl Polyp
28	85	6.5	635	22	AAE05378	Human Bcl-Gl Polyp
29	85	6.5	635	22	AAE05378	Human Bcl-Gl Polyp
30	83	6.3	331	21	AAE05378	Human Bcl-Gl Polyp
31	83	6.3	331	21	AAE05378	Human Bcl-Gl Polyp
32	83	6.3	331	21	AAE05378	Human Bcl-Gl Polyp
33	83	6.3	331	21	AAE05378	Human Bcl-Gl Polyp
34	82.5	6.3	1291	22	AAE05378	Human Bcl-Gl Polyp
35	82.5	6.3	1291	22	AAE05378	Human Bcl-Gl Polyp
36	81.5	6.2	250	21	AAE05378	Human Bcl-Gl Polyp
37	81.5	6.2	250	21	AAE05378	Human Bcl-Gl Polyp
38	81.5	6.2	250	21	AAE05378	Human Bcl-Gl Polyp
39	81.5	6.2	250	21	AAE05378	Human Bcl-Gl Polyp
40	81.5	6.2	250	21	AAE05378	Human Bcl-Gl Polyp
41	81.5	6.2	250	21	AAE05378	Human Bcl-Gl Polyp
42	81	6.2	807	20	AAE05378	Human Bcl-Gl Polyp
43	81	6.2	807	20	AAE05378	Human Bcl-Gl Polyp
44	80.5	6.1	777	19	AAE05378	Human Bcl-Gl Polyp
45	80.5	6.1	777	19	AAE05378	Human Bcl-Gl Polyp

ALIGNMENTS

RESULT 1	AAE07041	standard; Protein: 252 AA.
ID	AAE07041	standard; Protein: 252 AA.
AC	AAE07041	standard; Protein: 252 AA.
DT	23-OCT-2001	(first entry)
DE	Human BCL-X-like protein #2.	
XX	Human; BCL-X-like protein; therapy; physiological disorder.	
OS	Homo sapiens.	
XX	WO200157213-A2.	
XX	09-AUG-2001.	
XX	31-JAN-2001; 2001WO-US03446.	
XX	04-FEB-2000; 2000US-0180412.	
XX	(LEXI-) LEXICON GENETICS INC.	
XX	Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A, Zambrowicz B; Sands AT;	
XX	WPI: 2001-488882/53.	
XX	N-PSDB: AAD13336.	
XX	Novel isolated human BCL-X-like polynucleotide, useful in therapeutic, diagnostic and pharmacogenic applications	
XX	Claim 5; page 32; 33pp; English.	

XX The present sequence is human BCL-X-like protein.
CC The BCL-X-like polynucleotides are useful in therapeutic, diagnostic
CC and pharmacogenetic applications. They are useful for screening drugs
CC effective in the treatment of symptomatic or phenotypic manifestations
CC perturbing the normal function of protein in the body and also for
CC treating physiological disorders and diseases. The BCL-X-like
CC polynucleotides are useful in conjunction with polymerase chain
CC reaction to screen libraries, isolate clones, to prepare cloning
CC and sequencing templates and as hybridisation probes for assessing
CC gene expression patterns.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1316; DB 22; Length 252;
Best Local Similarity 100.0%; Pred. No. 9.7e-131;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCSTSGCDLEETPLDDDDLTETFEKILAYTRHHVFKSTPALFSPKILTRTSLSQKGLGN 60
F 1 mctsgcdleetpldddddltetfeklilaytrhhvfksfpalfspkiltrtslsqgqlgn 60
QY 61 CSANESWTEVSWPCRNOSSEKAINLGKSSWKAFEGVEKEKDSOSTPAKVSAOGQRTL 120
Db 61 csaneswtevwpcrnossekainlgksswkaffgvekedsgstpakvsagqgrtl 120
QY 121 EYODSHSQMSRCLSNVEQCLEHEAVDPKVIISANRYAEIYSWPPQATQAGGFSKKEI 180
Db 121 eyodshsqmsrclsnveqcleheavdpkvisianryaeiyswppqatqaggfkskel 180
QY 181 FVTEGLSFQLOGHVPVASSSKKDEEOILAKIVELLYSGDLEKRTAFIPPLVDTSI 240
Db 181 fvtelgsfqlqghvpassskkdeeqilakivellykysgdlerktafipplvdtsi 240
QY 241 QGFPODGLMACI 252
Db 241 qgfpgdglmaci 252

RESULT 2

AAE85167
ID AAE85167 standard; Protein; 252 AA.

XX AAE85167;
XX 07-SEP-2001 (first entry)
XX Human BCL-Gs polypeptide.
XX DE

XX BCL-Gs; cancer; cancer therapy; oncogene; apoptosis; BCL-Gs; cytostatic;
XX antiapoptotic; chromosome 12p12.3; human.
XX

XX Homo sapiens.
XX WO200144282-A2.
XX 21-JUN-2001.

XX 13-DEC-2000; 2000MO-US33793.
XX 14-DEC-1999; 99US-0461641.
XX (BURN-) BURNHAM INST.

XX Reed JC, Godzik A;
XX WPI: 2001-398125/42.
XX DR N-PSDB; AAH22583.

XX Novel polynucleotide encoding BCL-G polypeptide, useful for modulating
XX apoptosis, and for diagnosing and treating cancer -
XX

PS Claim 14; Fig 4; 117p; English.

XX The invention relates to BCL-G polypeptides and nucleic acids encoding
CC them. The BCL-G polypeptides can be expressed by standard recombinant
CC methodology. BCL-G oligonucleotides (or its anti-sense strand) and BCL-G
CC specific antibodies are useful for diagnosing cancer, monitoring cancer
CC therapy or assessing prognosis of patients with cancer. The BCL-G
CC polypeptides are useful for modulating the activity of an oncogenic
CC polypeptide. They are useful for identifying modulators, for modulating
CC a level of apoptosis mediated by the BCL-G polypeptide, for modulating
CC composition comprising the BCL-G polypeptide, polynucleotide or antibody
CC proliferation especially cancer. The present sequence represents a
CC human BCL-Gs polypeptide.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1316; DB 22; Length 252;
Best Local Similarity 100.0%; Pred. No. 9.7e-131;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MCSTSGCDLEETPLDDDDLTETFEKILAYTRHHVFKSTPALFSPKILTRTSLSQKGLGN 60
F 1 mctsgcdleetpldddddltetfeklilaytrhhvfksfpalfspkiltrtslsqgqlgn 60
QY 61 CSANESWTEVSWPCRNOSSEKAINLGKSSWKAFEGVEKEKDSOSTPAKVSAOGQRTL 120
Db 61 csaneswtevwpcrnossekainlgksswkaffgvekedsgstpakvsagqgrtl 120
QY 121 EYODSHSQMSRCLSNVEQCLEHEAVDPKVIISANRYAEIYSWPPQATQAGGFSKKEI 180
Db 121 eyodshsqmsrclsnveqcleheavdpkvisianryaeiyswppqatqaggfkskel 180
QY 181 FVTEGLSFQLOGHVPVASSSKKDEEOILAKIVELLYSGDLEKRTAFIPPLVDTSI 240
Db 181 fvtelgsfqlqghvpassskkdeeqilakivellykysgdlerktafipplvdtsi 240
QY 241 QGFPODGLMACI 252
Db 241 qgfpgdglmaci 252

RESULT 3

AAE07040
ID AAE07040 standard; Protein; 327 AA.

XX AAE07040;
XX 23-OCT-2001 (first entry)

XX Human BCL-X-like protein #1.
XX DE

XX Human; BCL-X-like protein; therapy; physiological disorder.
XX

XX Homo sapiens.
XX WO200157213-A2.

XX 09-AUG-2001.
XX 31-JAN-2001; 2001MO-US03446.

XX 04-FEB-2000; 2000US-0180412.
XX (LEXT-) LEXICON GENETICS INC.

XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A, Zambrowicz B;
XX Sands AT;
XX WPI: 2001-488882/53.
XX DR N-PSDB; AAD13235.
XX

PT Novel isolated human BCL-X-like polynucleotide, useful in therapeutic,
 PR diagnostic and pharmacogenic applications
 PS Claim 4; Page 30-31; 33pp; English.

XX The present sequence is human BCL-X-like protein.
 CC The BCL-X-like polynucleotides are useful in therapeutic, diagnostic
 CC and pharmacogenic applications. They are useful for screening drugs
 CC effective in the treatment of symptomatic or phenotypic manifestations
 CC perturbing the normal function of protein in the body and also for
 CC treating physiological disorders and diseases. The BCL-X-like
 CC polynucleotides are useful in conjunction with polymerase chain
 CC reaction to screen libraries, isolate clones, to prepare cloning
 CC and sequencing templates and as hybridisation probes for assessing
 CC gene expression patterns.

XX Sequence 327 AA;

Query Match 90.4%; Score 1189.5; DB 22; Length 327;
 Best Local Similarity 93.5%; Pred. No. 3.5e-117; Indels 7; Gaps 1;
 Matches 232; Conservative 2; Mismatches 7;

QY 1 MCSTSGCDLEERIPDDDDLNTIEFKILAYTRHHVFKSTPALESPKILTRSLSGRLGN 60
 1 mcsstsgcdleeripdddddntiefkllaytrhhvfkstpalfspklltrslsgrlgn 60
 DB 61 CSANESMTVEVSPKNSQSEKAINLGKRSKWAFFGVVEKEDSOSTPAKVSAGQRTL 120
 61 csanesmtvevspknsqsekainglkksswkaffgvvekedsgstpakvsagqrtl 120
 DB 121 EYODSHSQMSRCLSNVEQCLEHEAVDPKVISIANRVAEIVSWPPQATQAGFKSKEI 180
 121 eyodshsqmsrclsnveqcleheavdpkvisianrvaelyswppqatqagfkskel 180
 DB 121 eyqdsnsgwscrlsnveqcleheavdpkvisianrvaelyswppqatqagfkskel 180
 QY 181 FVTEGSLFQLOGHVPVASSSKKDEEOLIAKIVELKYSQDLEKRDIAFIPPLVDPSI 240
 181 fvtelgslfqloghvpvassskkdeeqllakivellkysgdqlerklk-----dkel 233
 DB 181 fvtelgslfqloghvpvassskkdeeqllakivellkysgdqlerklk-----dkel 233
 QY 241 QGFPODGL 248
 241 qgfpozgl 248
 DB 234 mghfdqgl 241

RESULT 4
 AAB85166 standard; Protein; 327 AA.

AAB85166;

XX 07-SEP-2001 (first entry)
 DT Human Bcl-G1 polypeptide.
 DE Human Bcl-G1 polypeptide.
 KW Bcl-G1; cancer; cancer therapy; oncogene; apoptosis; Bcl-G1; cytostatic;
 KW antiapoptotic; chromosome 12p12.3; human.
 OS Homo sapiens.
 OS Homo sapiens.
 PN WO200144282-A2.
 PD 21-JUN-2001.
 PF 13-DEC-2000; 2000MO-US33793.
 PR 14-DEC-1999; 99US-0461641.
 PA (BURN-) BURHAM INST.
 PI Reed JC, Godzik A;
 WPI: 2001-398125/42.
 DR N-PSDB; AAB85166.

XX Novel polynucleotide encoding Bcl-G polypeptide, useful for modulating
 PR apoptosis, and for diagnosing and treating cancer
 PS Claim 14; Fig 2; 117pp; English.

XX The invention relates to Bcl-G polypeptides and nucleic acids encoding
 CC them. The Bcl-G polypeptides can be expressed by standard recombinant
 CC methodology. Bcl-G oligonucleotides (or its anti-sense strand) and Bcl-G
 CC specific antibodies are useful for diagnosing cancer, monitoring cancer
 CC therapy or assessing prognosis of patients with cancer. The Bcl-G
 CC polypeptides are useful for modulating the activity of an oncogenic
 CC polypeptide. They are useful for identifying modulators for modulating
 CC a level of apoptosis mediated by the Bcl-G polypeptide. A therapeutic
 CC composition comprising the Bcl-G polypeptide, polynucleotide or antibody
 CC is useful for treating a pathology characterized by abnormal cell
 CC proliferation especially cancer. The present sequence represents a
 CC human Bcl-G1 polypeptide.

XX Sequence 327 AA;

Query Match 90.4%; Score 1189.5; DB 22; Length 327;
 Best Local Similarity 93.5%; Pred. No. 3.5e-117; Indels 7; Gaps 1;
 Matches 232; Conservative 2; Mismatches 7;

QY 1 MCSTSGCDLEERIPDDDDLNTIEFKILAYTRHHVFKSTPALESPKILTRSLSGRLGN 60
 1 mcsstsgcdleeripdddddntiefkllaytrhhvfkstpalfspklltrslsgrlgn 60
 DB 61 CSANESMTVEVSPKNSQSEKAINLGKRSKWAFFGVVEKEDSOSTPAKVSAGQRTL 120
 61 csanesmtvevspknsqsekainglkksswkaffgvvekedsgstpakvsagqrtl 120
 DB 121 EYODSHSQMSRCLSNVEQCLEHEAVDPKVISIANRVAEIVSWPPQATQAGFKSKEI 180
 121 eyodshsqmsrclsnveqcleheavdpkvisianrvaelyswppqatqagfkskel 180
 DB 121 eyqdsnsgwscrlsnveqcleheavdpkvisianrvaelyswppqatqagfkskel 180
 QY 181 FVTEGSLFQLOGHVPVASSSKKDEEOLIAKIVELKYSQDLEKRDIAFIPPLVDPSI 240
 181 fvtelgslfqloghvpvassskkdeeqllakivellkysgdqlerklk-----dkel 233
 DB 181 fvtelgslfqloghvpvassskkdeeqllakivellkysgdqlerklk-----dkel 233
 QY 241 QGFPODGL 248
 241 qgfpozgl 248
 DB 234 mghfdqgl 241

RESULT 5
 AAM95378 standard; Protein; 151 AA.

AAM95378;

XX 21-NOV-2001 (first entry)
 DT Human reproductive system related antigen SEQ ID NO: 4036.
 DE Human reproductive system related antigen SEQ ID NO: 4036.
 KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy.
 OS Homo sapiens.
 OS Homo sapiens.
 PN WO200155320-A2.
 PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01339.
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.

PR	13-OCT-2000	2000US-0239937.
PR	20-OCT-2000	2000US-0240960.
PR	20-OCT-2000	2000US-0241221.
PR	20-OCT-2000	2000US-0241785.
PR	20-OCT-2000	2000US-0241786.
PR	20-OCT-2000	2000US-0241787.
PR	20-OCT-2000	2000US-0241808.
PR	20-OCT-2000	2000US-0241809.
PR	01-NOV-2000	2000US-0241826.
PR	08-NOV-2000	2000US-0244617.
PR	08-NOV-2000	2000US-0246474.
PR	08-NOV-2000	2000US-0246475.
PR	08-NOV-2000	2000US-0246476.
PR	08-NOV-2000	2000US-0246477.
PR	08-NOV-2000	2000US-0246478.
PR	08-NOV-2000	2000US-0246523.
PR	08-NOV-2000	2000US-0246524.
PR	08-NOV-2000	2000US-0246525.
PR	08-NOV-2000	2000US-0246526.
PR	08-NOV-2000	2000US-0246527.
PR	08-NOV-2000	2000US-0246532.
PR	08-NOV-2000	2000US-0246538.
PR	08-NOV-2000	2000US-0246539.
PR	08-NOV-2000	2000US-0246610.
PR	08-NOV-2000	2000US-0246611.
PR	17-NOV-2000	2000US-0246613.
PR	17-NOV-2000	2000US-0249207.
PR	17-NOV-2000	2000US-0249208.
PR	17-NOV-2000	2000US-0249209.
PR	17-NOV-2000	2000US-0249210.
PR	17-NOV-2000	2000US-0249211.
PR	17-NOV-2000	2000US-0249212.
PR	17-NOV-2000	2000US-0249213.
PR	17-NOV-2000	2000US-0249214.
PR	17-NOV-2000	2000US-0249215.
PR	17-NOV-2000	2000US-0249216.
PR	17-NOV-2000	2000US-0249217.
PR	17-NOV-2000	2000US-0249218.
PR	17-NOV-2000	2000US-0249244.
PR	17-NOV-2000	2000US-0249245.
PR	17-NOV-2000	2000US-0249246.
PR	17-NOV-2000	2000US-0249265.
PR	17-NOV-2000	2000US-0249297.
PR	17-NOV-2000	2000US-0249299.
PR	01-DEC-2000	2000US-0249300.
PR	01-DEC-2000	2000US-0250160.
PR	05-DEC-2000	2000US-0250391.
PR	05-DEC-2000	2000US-0250393.
PR	05-DEC-2000	2000US-0250398.
PR	06-DEC-2000	2000US-0250719.
PR	08-DEC-2000	2000US-0251479.
PR	08-DEC-2000	2000US-0251856.
PR	08-DEC-2000	2000US-0251868.
PR	08-DEC-2000	2000US-0251869.
PR	08-DEC-2000	2000US-0251989.
PR	11-DEC-2000	2000US-0251990.
PR	05-JAN-2001	2000US-0254997.
PR		2000US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
DR	WPI; 2001-465570/50.	
XX	N-PSDB; AAL01348.	
XX	Isolated nucleic acid molecule encoding a reproductive system antigen	
PT	is used in preventing, treating or ameliorating a medical condition	
XX	Claim 11: SEQ ID NO 4036; 1297pp + Sequence Listing; English.	
PS	The present invention provides the protein and coding sequences of a	
CC	number of human reproductive system antigen.	

in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a protein of the invention.

Sequence 151 AA:

Query Match 52.5%; Score 691.5; DB 22; Length 151;
Best Local Similarity 95.0%; Pred. No. 5.6e-65;
Matches 134; Conservative 1; Mismatches 5; Indels 1; Gaps 1;

1 MCSTSGDLEIPLDDDDLTTEFKILAYTRHHVFKSPALFSPKILRTSLRSQGLGN 60
10 mcsstsgdleipldddddlttefkilaytrhhvfkstpalfspkiltrslrsqglgn 69
61 CSANESWTEVSWPCRNOSSEKAINIGKKSSWKAFFGVYEKEDSGSTPAKVSAGQRTL 120
70 csanewtevswpcrnossekainigksswkaffgvvekedsgstpakvsagqrtl 129
121 EYODSHSQWRCCLSNVEOCL 141
130 eyodshsqwrcclsnveocl 149

RESULT 6

ABB27961
ID ABB27961 standard; Peptide: 129 AA.

AC ABB27961;

01-FEB-2002 (first entry)

Human peptide #612 encoded by breast cell single exon nucleic acid probe.

Human: microarray; single exon probe; gene expression; breast;
disease; cancer.

Homo sapiens.

WO200157271-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00662.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0633366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes,
useful for measuring gene expression in sample derived from human
breast, comprises number of single exon nucleic acid probes -

Claim 27; SEQ ID NO 10929; 327pp + sequence listing; English.

The invention relates to a spatially-addressable set of single exon
nucleic acid probes for measuring gene expression in a sample derived
from human breast and BT 474 cells. The method involves contacting
the probes with a collection of detectably labelled nucleic acids
derived from mRNA of human breast, and then measuring the label
bound to each probe of the microarray. The probes are useful for
verifying the expression of regions of genomic DNA predicted to
encode proteins. They are useful for gene discovery, and for
determining predisposition and/or prognostic breast disease. Gene

expression analysis is useful for assessing the toxicity of chemical
agents on cells. The microarray of this invention presents a far greater
diversity of probes for measuring gene expression, with far less bias
than expressed sequence tag microarrays. The method is suitable for
rapid production of functional information from genomic sequence. The
present sequence is a peptide encoded by a single exon nucleic acid
probe of the invention.
Note: The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pcr_sequences.

Sequence 129 AA:

Query Match 52.1%; Score 686; DB 22; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.7e-64;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 DDDLTTEFKILAYTRHHVFKSPALFSPKILRTSLRSQGLGNCSANESWTEVSWPCR 75
1 dddlttefkilaytrhhvfkstpalfspkiltrslrsqglgncsanewtevswpcr 60
76 NSQSEKAINIGKKSSWKAFFGVYEKEDSGSTPAKVSAGQRTLEYODSHSQWRCCL 135
61 nsqsekainigksswkaffgvvekedsgstpakvsagqrtleyodshsqwrccls 120
136 NVEOCLHE 144
121 nveoclhe 129

RESULT 7

ABB33133
ID ABB33133 standard; Peptide: 129 AA.

AC ABB33133;

04-FEB-2002 (first entry)

Peptide #639 encoded by human foetal liver single exon probe.

Human: foetal liver; gene expression; single exon nucleic acid probe.

Homo sapiens.

WO200157277-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00669.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0633366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human foetal liver -

Claim 27; SEQ ID NO 25768; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for
measuring human gene expression in a sample derived from human foetal
liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human
 CC fetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 129 AA;

Query Match 52.1%; Score 686; DB 22; Length 129;
 Best Local Similarity 100.0%; Pred. No. 1.7e-64;
 Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 DDDLTIEFKILAYTRHHVFKSTPALFSPKILRTSLISQRLGNCANESWTVEWPCR 75
 Db 1 dddltiefkilaaytrhhvfkstpalfspkllrtslsqrglncanestevswpcr 60
 QY 76 NSOSSEKAINLCKKSSWKAFFGVVEKEDSOSTPAKVSAGQRTLEYQDSHSQWSRCLS 135
 Db 61 nsqsekainlgkkskswkaffgvvekedsgstpakvsagqrileyqdsqswsrcls 120
 QY 136 NVEQCLEHE 144
 Db 121 nveqclehe 129

RESULT 8

ABBI8598
 ID ABB18598 standard; Protein: 129 AA.
 AC ABB18598;

DT 23-JAN-2002 (first entry)
 DE Protein #597 encoded by probe for measuring heart cell gene expression.

XX Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KM congenital heart disease.
 XX Homo sapiens.

PN WO200157274-A2.
 PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00666.
 XX

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-48899/53.

PT Single exon nucleic acid probes for analyzing gene expression in human
 XX hearts -

PS Claim 15; SEQ ID No 20368; 530pp; English.
 XX

CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA153-ABA41305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,
 CC diagnosing, grading, staging, monitoring and proposing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 129 AA;

Query Match 52.1%; Score 686; DB 22; Length 129;
 Best Local Similarity 100.0%; Pred. No. 1.7e-64;
 Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 DDDLTIEFKILAYTRHHVFKSTPALFSPKILRTSLISQRLGNCANESWTVEWPCR 75
 Db 1 dddltiefkilaaytrhhvfkstpalfspkllrtslsqrglncanestevswpcr 60
 QY 76 NSOSSEKAINLCKKSSWKAFFGVVEKEDSOSTPAKVSAGQRTLEYQDSHSQWSRCLS 135
 Db 61 nsqsekainlgkkskswkaffgvvekedsgstpakvsagqrileyqdsqswsrcls 120
 QY 136 NVEQCLEHE 144
 Db 121 nveqclehe 129

RESULT 9

AAM53929
 ID AAM53929 standard; Protein: 129 AA.
 AC AAM53929;

DT 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26034.

XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KM epilepsy; cancer.
 XX Homo sapiens.

PN WO200157275-A2.
 PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.
 XX

PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-48346/52.

PT Single exon nucleic acid probes for analyzing gene expression in human
 XX brains -

PS Example 4; SEQ ID NO: 26034; 650pp + Sequence Listing; English.
 XX

CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX

SO Sequence 129 AA:

Query Match 52.1%; Score 686; DB 22; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.7e-64;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 DDDNTIEFKIIAYYTRHHVFKSTPALFSPKLLRTRSLQSGNGCSANESWTEVSPCR 75
Db 1 dddntiefkIIayYtrhhvfksTPALfSPKllrtrslsqrglncsaneswtevsWpcr 60
QY 76 NSQSEKAINLGKKSSWKAFFGVVEKEDSOSTPAKVSAGORTLEYODSHSQMSRCLS 135
Db 61 nsqsekainlgkkskwkaffgvvekedsgstpakvsagqrileyqdsnsgqwsrcls 120
136 NVEQCLEHE 144
121 nveqclehe 129

RESULT 10
AAM6317
ID AAM6317 standard; Protein: 129 AA.

AC AAM6317;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 26623.
DE Human bone marrow expressed exon; gene expression analysis; probe;
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukemia; lymphoma; myeloma.
XX Homo sapiens.

OS Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI: 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX Example 4; SEQ ID NO: 26623; 658bp + sequence listing; English.

XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX Sequence 129 AA;

Query Match 52.1%; Score 686; DB 22; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.7e-64;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 DDDNTIEFKIIAYYTRHHVFKSTPALFSPKLLRTRSLQSGNGCSANESWTEVSPCR 75
Db 1 dddntiefkIIayYtrhhvfksTPALfSPKllrtrslsqrglncsaneswtevsWpcr 60
QY 76 NSQSEKAINLGKKSSWKAFFGVVEKEDSOSTPAKVSAGORTLEYODSHSQMSRCLS 135
Db 61 nsqsekainlgkkskwkaffgvvekedsgstpakvsagqrileyqdsnsgqwsrcls 120
136 NVEQCLEHE 144
121 nveqclehe 129

RESULT 11
AAM14186
ID AAM14186 standard; Protein: 129 AA.

AC AAM14186;
XX 12-OCT-2001 (first entry)
XX Peptide #620 encoded by probe for measuring cervical gene expression.
DE Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX Homo sapiens.

OS Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00670.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI: 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX Claim 27; SEQ ID No 19012; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes
XX (SENP: see A110068-A128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 129 AA;

Db 1 dddlnlefkllaylrhvfksfpal fspkllrtrlsqrglgnscaneswvewmpcr 60
 QY 76 NSSSKKAINLGKKSSKMAFEVGEKEDSQSNPAKVSAGQRTLEYODSHSOQMSFCLS 135
 Db 61 nsssskainlgkkskswkafgvekedsqspakvsagqgrtlleyqdashsqwscrls 120
 QY 136 NVEQCLEHE 144
 Db 121 nveqclehe 129
 RESULT 14
 AAB85188 standard; protein; 328 AA.
 AAB85188;
 07-SEP-2001 (first entry)
 Mouse Bcl-G polypeptide.
 Bcl-G; cancer; cancer therapy; oncogene; apoptosis; cytostatic;
 antiapoptotic; chromosome 12p12.3; mouse.
 Mus sp.
 Key Location/Qualifiers
 Misc-difference 107
 /label= unknown
 /note= "encoded by NCG"
 W0200144282-A2.
 21-JUN-2001.
 13-DEC-2000; 2000MO-US33793.
 14-DEC-1999; 99US-0461641.
 (BURN-) BURNHAM INST.
 Reed JC, Godzik A;
 WPI: 2001-398125/42.
 DR N-PSDB; AAB22670.
 Novel polynucleotide encoding Bcl-G polypeptide, useful for modulating
 apoptosis, and for diagnosing and treating cancer -
 Claim 14; Page 114-115; 117pp; English.
 The invention relates to Bcl-G polypeptides and nucleic acids encoding
 them. The Bcl-G polypeptides can be expressed by standard recombinant
 methodology. Bcl-G oligonucleotides (or its anti-sense strand) and Bcl-G
 specific antibodies are useful for diagnosing cancer, monitoring cancer
 therapy or assessing prognosis of patients with cancer. The Bcl-G
 polypeptides are useful for modulating the activity of an oncogenic
 polypeptide. They are useful for identifying modulators for modulating
 a level of apoptosis mediated by the Bcl-G polypeptide. A therapeutic
 composition comprising the Bcl-G polypeptide, polynucleotide or antibody
 is useful for treating a pathology characterized by abnormal cell
 proliferation especially cancer. The present sequence represents a
 mouse Bcl-G polypeptide.
 Sequence 328 AA:
 Query Match 38.1%; Score 501; DB 22; Length 328;
 Best Local Similarity 47.5%; Pred. No. 2.5e-44;
 Matches 122; Conservative 33; Mismatches 78; Indels 24; Gaps 8;
 1 MCSTSGCDLEIPLDDDLNIEFKILAYTRRHVFKSPALFSPKLLRTRLSQSGGLGN 60

Db 1 mcsstvydledipreddpnstiekliafayrhvfkntpavfapkslstrlsqkalgt 60
 QY 61 CSANESWTEVSWPCRNSSQSEKAINLCKKSSKMAFEVGEKEDS-QSTAPKVSAGQCR- 118
 Db 61 ws-ldswtqyslporgspssknslgkkswtllfrveakeeglpxspketragspg 119
 QY 119 --TLEYODS-HSQQMSRCLSNVEQCLEHEAVDPKVISIANRVAEIVYSWPPQATQAGCF 175
 Db 120 pfpyergsgfngdmprslsveqpwrvklypkwlypewklf---tpghqmsst 175
 QY 176 KSKELFVTEG---LSEQLQGHVPAVSSSKRDEEQQIAKIVELIKYSGDQLEKDTAFI 231
 Db 176 areeasskrgrfflyffteq---pwdsknkdgedqllskivellkssgddlgreikx-- 230
 QY 232 PIPLYTISIQGPPODGL 248
 Db 231 ----dkalmssfdqgl 242
 RESULT 15
 AAG65914 standard; protein; 1250 AA.
 AAG65914;
 11-FEB-2002 (first entry)
 Amino acid sequence of GSK gene id 27142.
 Peptide hormone; antidiabetic; anorectic; antianorectic; antiasthmatic;
 antidepressant; nootropic; neuroprotectant; hypotensive; hypertensive;
 cytostatic; cerebroprotective; vasotropic; human.
 Homo sapiens.
 W0200172961-A2.
 04-OCT-2001.
 22-MAR-2001; 2001MO-US92226.
 24-MAR-2000; 2000US-192158P.
 28-MAR-2000; 2000US-192668P.
 27-APR-2000; 2000US-200166P.
 (SMIK) SMITHKLINE BEECHAM CORP.
 (SMIK) SMITHKLINE BEECHAM PLC.
 Agarwal P, Murdoch PR, Rizvi SK, Smith RF, Xiang Z, Rabnick KS;
 Lal Y;
 WPI: 2001-639223/73.
 DR N-PSDB; AAI67204.
 Isolated polypeptides, which may be peptide hormones, which are
 identified by high throughput genome-based biology which identifies
 genes and gene products as therapeutic targets for treatment of
 diseases such as diabetes and cancer -
 Claim 1; Page 91-94; 99pp; English.
 The invention provides polypeptides (AAG65886-65918) which may be peptide
 hormones (including insulin, growth hormones, chemokines, cytokines,
 neuropeptides, integrins, kallikreins, lamins, melanins, nutritive
 hormones, neuropepsin, selectins, thromboglobulins, pleiotrophins, prostaglandins,
 secretogranins, thymosins, thymosins) identified by
 high throughput genome-based biology and polynucleotides (AAI67176-67208)
 encoding them. The polypeptides can be expressed by standard recombinant
 methodology. The polypeptides are useful in the treatment of disease such
 as diabetes, breast-, prostate-, colon cancer and other malignant tumors,
 hyper- and hypotension, obesity, bulimia, anorexia, growth abnormalities,
 asthma, manic depression, dementia, delirium, mental retardation,


```
BASE COUNT      289 a      230 c      250 g      215 t
ORIGIN          /organism="Homo sapiens"
                /db_xref="taxon:9606"
```

Query Match 100.0%; Score 984; DB 6; Length 984;
 Best Local Similarity 100.0%; Pred. No. 5.8e-271;
 Matches 984; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 atgttaagacacagtggtgtgacgtggaagaatccccctagatgatgaactaac 60
    |||||||
DB 1 ATGTGTAGACACAGTGGGTGTGACCTGGAGAAATCCCCCTAGATGATGACTTAAC 60
OY 61 accatagaatccaanaatcctgcctactacacacagacatcatgtctccaagacccct 120
    |||||||
DB 61 ACCATAGAAATCAAAATCCCTGCCTACTACACAGACATCATGTCTTCAAGACACCCCT 120
OY 121 gctctcttccacaaagctgctgtagaagaagtgttcccaagaggggcttggagat 180
    |||||||
DB 121 GCTCTCTTCTCAACCAAGCTGTGAGAACAGAGTTGTGCCAGAGGGGCTGGGAGAT 180
OY 181 tttcaagcaaatgagtcagagagagtgatgagccttgcaagaatcccaatccagt 240
    |||||||
DB 181 TCTTCAGCAAAATGAGTCATGACAGAGGTGTCAATGCGCTTCACAAATCCCAATCCAGT 240
OY 241 gagaaagccataaaccttggcaagaanaagtcctcttgaagaagatcttggagtagtg 300
    |||||||
DB 241 GAGAAAGCCATAAACCTTGSCAAGAAAGTCTTGGAAAGCATTTCTTGGAGTAGTG 300
OY 301 gagaaaggaagatctgcagagcagcgcctgccaaggtctctgctcaggggtccaagagcgttg 360
    |||||||
DB 301 GAGAAAGGAAGATTGCGACAGACAGCCTGCGCAAGGTCTCTGCTCAGGGTCAAGAGACGTTG 360
OY 361 gataccaagatctgcacagccagcagtggttccagtgctcttccaaagtgtagcagtg 420
    |||||||
DB 361 GAAATCAAGATTCGACACAGCAGCAGTGGTCCAGGGTCTTTTCAACGTGACAGCATGCC 420
OY 421 ttggagatgaagctgttgagaccccaagtaattccattcccaacccagtagctgaatt 480
    |||||||
DB 421 TTGGAGCATGAAGCTGTGAGACCCCAAGTCAATTCATTCGCAACCGAGTACTGAAT 480
OY 481 gtttattcttgcacacacacacagcagcagcagtaggcttcaagctccaagaagatt 540
    |||||||
DB 481 GTTTATTCTTGGCCACACCCACACAGCCAGGAGGAGGCTTCAATCCCAAGAGATT 540
OY 541 ttgttaactgaaggtctctctctcccaagctccaagcagctgcttgaagctcaagttct 600
    |||||||
DB 541 TTTGTAACTGAGGGTCTCTCTTCCAGCTCCAAAGCCAGTGGCTTGAAGCTTCAAGTTCT 600
OY 601 aagaaagatgaagaagaacaataactagccaaatgtttagagctgctgaataatccaaga 660
    |||||||
DB 601 AAGAAAGATGAAGAAAGAACTAAGCAAAATGTGTGAGCTGCTGAAAAATTCAGGA 660
OY 661 gatcaattggaagaagaagctgaagaagaataaggtcttgaatggagccacttccagatggg 720
    |||||||
DB 661 GATCACTTGGAAAGAAAGCTGAAAGAAAGATTAAGCTTTGATGGGCACTTCCAGATGGG 720
OY 721 ctgtcctactctgttcttcaagacatcacagacagctcctaattggtgtggaacccag 780
    |||||||
DB 721 CTGTCTACTCTGTTTCAAGACCATCACAGACAGGTCCTAATGGGTGTGAGACCCAG 780
OY 781 ggaagaatcagaggtcaagagctcagaggtttaaagctgctctgttaatagactcaagggc 840
    |||||||
DB 781 GGAGAAATCAGAGGTCAAAAGCTCAAGGCTTTAAGGCTGCCCTTGAATGACCTCAGG 840
OY 841 aagctcagagctatgacaacacacccgatgaagaaggtcctggagcttgggcaagaatgc 900
    |||||||
DB 841 AAGCTCAGACCTTTGACAAACACCCGATGAACAGGTCCTGGGCTTTGGCACCAAGTAC 900
OY 901 ctgaagaagaactcttcgcatggatccagcagcagcttggatgggaaanaaatacttggg 960
    |||||||
DB 901 CTGAAGAGAACTTCTGCGCATGATCAGCAGCAGCTGTGATGGAAAAAATACTTGGG 960
OY 961 atatacatgaagaagaagtagctga 984
    |||||||
DB 961 ATATCACATGAAGAGTAGACTGA 984
  
```

RESULT 2
 AX174783
 LOCUS
 DEFINITION
 AX174783
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 1179)
 AUTHORS
 TITLE
 JOURNAL
 REED, J.C. and Godzik, A.
 Bcl-2 polypeptides: encoding nucleic acids and methods of use
 Patent: WO 0144282-A 1 21-JUN-2001;
 The Burnham Institute (US)
 location/Qualifiers
 1..1179
 /source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 196..1179
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC43798.1"
 /db_xref="GI:14598300"
 /translation="MCSRSQDLEIPRDDDIANTIEFKILAYTRRHVFKSPALF
 PKILRTSLRSLRGICNSANESWTEVSPKRNSSOSEKAINIGKKSSMKARFVYEK
 EDSOSTPAKVSAGQRTLEYDSDSHQNSRCLSNVECLHEAVDPKVLSTIANRAEI
 VYSPMPQAOVQAGGFKSELEIVTEGLSPQLGHNVAASSKDEDEQLIAKLIVELKY
 SDQLERLKKDKALMGHFDGLSYVFKITTDVLYMGVDPGESEVKNQGRKALVI
 DVTAKLTAIINHPNRYLGFSTKYLKNEFSWIOHGGMEKILGISHEVD"

BASE COUNT 340 a 285 c 305 g 249 t
 ORIGIN

Query Match 99.8%; Score 982.4; DB 6; Length 1179;
 Best Local Similarity 99.9%; Pred. No. 1.7e-270;
 Matches 983; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 atgtgtgacacagtggtgtgacgtggaagaatccccctagatgatgaactaac 60
    |||||||
DB 196 ATGTGTAGACACAGTGGGTGTGACCTGGAGAAATCCCCCTAGATGATGACTTAAC 255
OY 61 accatagaatccaanaatcctgcctactacacacagacatcatgtctccaagacccct 120
    |||||||
DB 61 ACCATAGAAATCAAAATCCCTGCCTACTACACAGACATCATGTCTTCAAGACACCCCT 315
OY 256 accatagaatccaanaatcctgcctactacacacagacatcatgtctccaagacccct 120
    |||||||
DB 256 ACCATAGAAATCAAAATCCCTGCCTACTACACAGACATCATGTCTTCAAGACACCCCT 315
OY 121 gctctcttccacaaagctgctgtagaagaagtgttcccaagaggggcttggagat 180
    |||||||
DB 121 GCTCTCTTCTCAACCAAGCTGTGAGAACAGAGTTGTGCCAGAGGGGCTGGGAGAT 375
OY 316 gctctcttccacaaagctgctgtagaagaagtgttcccaagaggggcttggagat 180
    |||||||
DB 316 GCTCTCTTCTCAACCAAGCTGTGAGAACAGAGTTGTGCCAGAGGGGCTGGGAGAT 375
OY 181 tttcaagcaaatgagtcagagagagtgatgagccttgcaagaatcccaatccagt 240
    |||||||
DB 181 TTTCAAGCAAAATGAGTCATGACAGAGGTGTCAATTCGCAACCGAGTACTGAAT 435
OY 241 gagaagaagcctaagaacttggcagaagaagaagcttcttggaaagcatttcttgaagtagtg 300
    |||||||
DB 241 GTTTCAGCAAAATGAGTCATGACAGAGGTGTCAATTCGCAAAATTCCTCAATCCAGT 435
OY 436 gagaagggccatnaaaccttggcagaanaaagcttcttggaaagcatttcttgaagtagtg 300
    |||||||
DB 436 GAGAAAGCCATNAACCTTGGCAGAAAGAGTCTTCTTGGAAAGCATTTCTTGGAGTAGTG 495
OY 301 gagaagaagaatcgcaagacagcctgccaagctctctgctcagaggtccaagaagcttg 360
    |||||||
DB 301 GAGAAGGAAGATTGCGACAGACAGCCTGCGCAAGGTCTGTGCTCAGGCTCAAGAGAGCTTG 555
OY 496 gagaagaagaatcgcaagacagcagtggttccagtgctcttctaagtgagagctg 420
    |||||||
DB 496 GAGAAGGAAGATTGCGACAGACAGCCTGCGCAAGGTCTGTGCTCAGGCTCAAGAGAGCTTG 555
OY 361 gaataccaagatctgcacagcagcagtggttccagtgctcttctaagtgagagctg 420
    |||||||
DB 361 GAAATCAAGATTGCGACAGCAGCAGTGGTCTTCTTCAACGTGAGCAAGTGC 615
OY 556 gaaatcaagaatttgcacacagcagcagtggttccagtgctcttctaagtgagagctg 420
    |||||||
DB 556 GAAATCAAGATTGCGACAGCAGCAGTGGTCTTCTTCAACGTGAGCAAGTGC 615
OY 421 ttgagacatgaagctgttgaccccaagatcatttccattggccaacagagtagctgaatt 480
    |||||||
DB 421 TTGAGACATGAAGCTGTGAGACCCCAAGATCAATTTCCATTGGCCAACGAGTAGCTGAAT 675
  
```


Df	301	GAGAAAGGAGATTTGGCAAGACACGCCCTCCAAAGGTCTCTGCCTCAGGTCFCAAAGACAGTTC	360
Oy	361	gaataccaagattggcaaacgacgaagatggtccaaagtcttcttaacagtggagaatgc	420
Df	361	GAATTACCAGATTTGGCCACACGACGACGACGATGGTGCCAGGTCTTTTAAGCTGAGCATGC	420
Oy	421	ttagcagcatgaagctgttgaccaccaaatcatlctcatcttcgaaccgagtagctaaatt	480
Df	421	TTCGAGCATGAAGCTGTGTGACCCCAAAATCTTTTCCATTGSCAACCGAGTAGCTAAATT	480
Oy	481	gttatctccgtgccacaaccaacgaagcgagaccaggcagaagcttcaaagataatt	540
Df	481	GTTTATTCTTCGGCCACACCAACGCGACCGAGGAGAGGCTTCAACTCCAAGAAGATTT	540
Oy	541	tttctaactggaggtctctctcttcacagtlccaagcgacgtgcgtgatctcaattct	600
Df	541	TTTTGTAATGAGGCTCTCTCTTCCTTCAGCTCCAAAGGCCACGTGCTGTGCTTCAAGTTCT	600
Oy	601	aagaagaatgaagaagaacaatactagccaanaattgttagcgtgctgaatatatcaga	660
Df	601	AAGAAAGATGAAAGAAAGACAATACTACCCAAAATTGTTGAGCTGCTGAATAATTCACAGA	660
Oy	661	gatacgttggaagaagaag 678	
Df	661		
RESULT	7		
LOCUS	AY040274	1176 bp	mRNA linear PRI 01-AUG-2001
DEFINITION	Homo sapiens apoptosis regulator BCL-G median form (BCLG) mRNA,		
ACCESSION	AY040274		
VERSION	AY040274.1	GI:15072494	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 1176)		
TITLE	Montpetit,A., Boily,G. and Sinnett,D.		
JOURNAL	A detailed transcriptional map of the chromosome 12p12 tumor		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1176)		
TITLE	Montpetit,A. and Sinnett,D.		
JOURNAL	Direct Submision		
REFERENCE	Submitted (14-JUN-2001) Hemato-Oncology, Hospital Ste-Justine, 3175		
AUTHORS	Cote-St-Catherine, Montreal, QC H3T 1C5, Canada		
TITLE	Location/Qualifiers		
SOURCE	1..1176		
gene	/organism="Homo sapiens"		
CDS	/db_xref="taxon:9606"		
	/chromosome="12"		
	/map="12p12"		
	1..1176		
	/gene="BCLG"		
	8..838		
	/gene="BCLG"		
	/note="member of the BCL-2 family"		
	/codon_start=1		
	/product="apoptosis regulator BCL-G median form"		
	/protein_id="AAK2109.1"		
	/db_xref="GI:15072495"		
	/translation="WCSTSGCDLEELPLDDDDINTIEFKILAYTRHHVKSPPALFS		
	PKLIRTSLSQRLGNGSANEWSWTEVSWPKNSSSKAINTLGRKKSSWRAPFGVEEK		
	EDQSRTFAKVAAGORTLEODSHSQOOSRCLSNVEQCLEHAEADPKVISITANKVAEI		
	VYSMPDQAOTAGGFKEIFEVGLGSFOLOGHPVASSSSKKDDEEOIILAKIVIELKY		
	SGQLRKKEKMEESAKTKIKMLIMSSSGGPRVLISIPAFKNNSTLTLPPESEHN		
BASE COUNT	363 a	277 c	270 g 266 t
ORIGIN			

Query Match	Similarity	Score	DB	Length
Best Local	98.6%	Pred. No. 3.2e-183		
Matches	684	Conservative	0	Mismatches 10; Indels 0; Gaps 0;

LOCUS	DEFINITION	SEQUENCE	FROM	PATENT	WU0144282.1	GI:14598301
AX174785	Sequence 3 from Patent WU0144282.1	AX174785	954 bp	DNA	Linear	PAT 03-JUL-2001
VERSION	AX174785.1	GI:14598301				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 954)					
TITLE	Reed, J.C. and Godzik, A.					
JOURNAL	Bcl-2 polypeptides, encoding nucleic acids and methods of use					
FEATURES	Patent: WO 0144282-A 3 21-JUN-2001;					
SOURCE	The Burnham Institute (US)					
	Location/Qualifiers					
	1..954					

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE
Db	537	ACCCACTCAAAAGAGAGGCTCCGAGATTTTGTACTTCAGGTTTGAAGACCTTGGA	596			
Qy	586	gtaagctcaagctcctaagaagaatgaaagaagaacaaatctagccaaatctgttgactg	645			
Db	597	CT-----CTAAGAATTAAGATGCTTAAGACCAAAATATATAAGCAAGATGTGACCTG	648			
Qy	646	ctgaatatcagagatcaagcttgaaagaagcctgaagaagaataagcttgatggc	705			
Db	649	CTGAATTCCTCGGGGGGACCTTGGGAAGAGAGATTAAGAAAGACCAAGCCTTTCATGACG	708			
Qy	706	caactcagaagatgggctgtccctactctgttttcaaagaccatacagaaccagctctaag	765			
Db	709	AGCTTCACAGACGGGGCTGTCTTACTCAACGTTCAAGACCATCACAGACCTGTTCCTGAGG	768			
Qy	766	ggtgtgagaccccaaggagaaatcagagtcgaagtcgaagcttaagcttgcccttgta	825			
Db	769	GACGTGGACACCCAGAGAGAGAAATCAGAGATCAAAAGCTCGGGGCTTCAAGGCTGCCCTTGCA	828			
Qy	826	atagaagctacagcccaagctcacagctatgtgaacaaccaccgaatgaacaggtccctggc	885			
Db	829	ATAGAGGCCATTCGCCAAGCTCACGGCATCGGACAAACACCCATGATAGAAATGCTGGGC	888			
Qy	886	tttgaccacaagctacctcgaagaagaactcttcgccatgagatccagcagcaggttgatg	945			
Db	889	TTGCGGACCAAGTACCTAAAGAGACTTCTCCCTCCCTGGGTTGACAGAGATGGGGATGG	948			
Qy	946	gaaaaaatacttggtgatatcacatgataagaagtaactga	984			
Db	949	GAAAAAATACTTGGGATCTCATAGAAAGATGACTGA	987			
RESULT 10						
LOCUS	AC007537	190858 bp	DNA	linear	PRI 09-OCT-1999	
DEFINITION	Homo sapiens 12p12 BAC RPi11-267J23 (Roswell Park Cancer Institute					
ACCESSION	AC007537					
VERSION	AC007537.3	GI:4914348				
KEYWORDS	HTG.					
ORGANISM	human.					
SOURCE	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.					
1 (bases 1 to 190858)						
Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,						
Dugan,S., Durbin,U., Forcum,U., Garcia,C., Gorell,J.H.,						
Gorell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,						
Kondelwsk,N., Lau,S., Leal,B., Lee,E., Lichtenarge,O., Liu,W.,						
Logan,D., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,						
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,						
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,						
Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W.,						
Zhou,X., Kuchelapatti,R., Nelson,D. and Gibbs,R.A.						
Direct Submission						
Unpublished						
2 (bases 1 to 190858)						
Worley,K.C.						
Direct Submission						
Submitted (13-MAY-1999)						
Molecular and Human Genetics, Baylor						
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
3 (bases 1 to 190858)						
Worley,K.C.						
Direct Submission						
Submitted (27-MAY-1999)						
Molecular and Human Genetics, Baylor						
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
4 (bases 1 to 190858)						
Worley,K.C.						
Direct Submission						
Submitted (28-MAY-1999)						
Human Genome Sequencing Center, Department						
of Molecular and Human Genetics, Baylor College of Medicine, One						
Baylor Plaza, Houston, TX 77030, USA						
5 (bases 1 to 190858)						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						

AUTHORS
TITLE
JOURNAL

Worley, K.C.
Direct Submission
Submitted (09-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 28, 1999 this sequence version replaced gi:4895156.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
sc-hell@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSS are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats were identified using RepeatMasker (unpublished) for human and mouse sequences. Genes and regions of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Config length: 190858
Phrap values in estimate: 188824
Average error rate (BCM-Phrap estimate): 0.00023902
Fraction of Phrap values less than 40 : 0.039746
Number of consensus changing edits: 12
Number of N's in consensus : 0
```

	Consensus	Changing	edits
Position	Original+Context	Edited+Context	
16166	atagtgtgaac(c)ccctactctc	atagtgtgaac(a)ccctactctc	
26355	taatttttca(n)acaggcaaaa	taatttttca(a)acaggcaaaa	
35016	ttttttttt(n)aggcgaaagtc	ttttttttt(g)aggcgaaagtc	
35923	ttattattat(n)ttttttttt	ttattattat(t)ttttttttt	
49346	agccagtcgga(n)cttgcacgtga	agccagtcgga(g)cttgcacgtga	
49663	tggcccaatc(n)gtgtaaaacc	tggcccaatc(a)gtgtaaaacc	
80511	gagccggag(n)ttgacacagag	gagccggag(c)ttgacacagag	
91836	gaagaataa(n)accagctctg	gaagaataa(a)accagctctg	
9042728	accatttag(t)tagtagacag	accatttag(t)tagtagacag	
114732	ctccctccc(n)ccctccc	ctccctccc(c)ccctccc	
118543	cattgttgaac(n)cccgctctca	cattgttgaac(a)cccgctctca	
189415	gattcttca(n)cccaggagag	gattcttca(a)cccaggagag	

----- Distribution of Quality < 40 Bases

Db 177088 GCTCTCTTCTACCAAACTGCTGAGACAGAGTTTGTCCAGAGGGCCCTGGGGAAT 177029
Qy 181 Tgttcagcaaatgagtcagtcagagagtgctcatgagccttcagaaatcccaatcagc 240
Db 177028 TGTTCAGCAAAATGAGTCATNGACAGAGTGTCATGCGCTTTCGAGAAATTCACATCAGT 176969
Qy 241 gagaagcgcataaacttcgcaagacgccttcgcaagaaagtccttcgtgaaagcattcctttgagtagtg 300
Db 176968 GAGAAGGCCTAAACCTTGGCAGAAAGTCTTCTTGGAAACATCTTTCGAGTAGTG 176909
Qy 301 gagaaggaagatctgcgaagacgccttcgcaagagtccttcgtcagtaggtcaagagcagcttg 360
Db 176908 GAGAAGGAGAGTATTCGACAGACGCGCTGCCAAGGTCTGCTCAGGTCAAGAGACGTTG 176849
Qy 361 gaatacaagaatctgcgaagcagcagtcagtcagagtccttcctaagctggaagcagctgc 420
Db 176848 GAATACCAAGATTTGCGACAGCAGCAGTGTCAGGTGCTTCTTAACGTGAGACATGTC 176789
421 ttggagcatgaagctgtggaaccccaagtcatttc 455
Db 176788 TTGGAGCATGAGAGTGAGCATCTGGGATTTCTTTC 176754
RESULT 12
AC091816 200499 bp DNA linear HTG 02-OCT-2001
LOCUS Homo sapiens chromosome 12p clone RP11-52513, WORKING DRAFT
DEFINITION AC091816
ACCESSION AC091816
VERSION AC091816.7 GI:15625994
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULFILLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 200499)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barberia,J.,
Banton,J., Bimagne,K., Blankenburg,K., Bonin,D., Bouck,J.,
Bowle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douhwaite,K.J., Draper,H.,
Dugan-Rocha,S., Dublin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gall,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homs,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivel,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,J.C.,
Lewis,L., Li,J., Li,Z., Litcharge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegue,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenko,S.,
Ogih,M., Okwono,G., Orangunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubokan,I., Rolfe,M.,
Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Swatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlecczyk,R., Wooden,S.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 200499)
Worley,K.C.
Direct Submission
Submitted (09-JUN-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 16, 2001 this sequence version replaced gi:15042773.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HDKE
Center clone name: RP11-52513
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 20651 bases at least Q40
Consensus quality: 207841 bases at least Q30
Consensus quality: 208973 bases at least Q20
Estimated insert size: 203052; sum-of-coverage estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 12.7x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 48113: contig of 48113 bp in length
* 48114: gap of unknown length
* 48214: gap of 37737 bp in length
* 85951: gap of unknown length
* 85952: gap of unknown length
* 110963: contig of 24913 bp in length
* 110964: gap of unknown length
* 110965: gap of unknown length
* 111064: contig of 21162 bp in length
* 132226: gap of unknown length
* 132227: gap of unknown length
* 132326: contig of 14879 bp in length
* 147205: gap of unknown length
* 147206: contig of 14742 bp in length
* 162047: gap of unknown length
* 162147: gap of unknown length
* 162148: gap of unknown length
* 176916: gap of unknown length
* 177015: gap of unknown length
* 189651: contig of 12636 bp in length
* 189652: gap of unknown length
* 189752: 200499: contig of 10748 bp in length.
*
FEATURES
source
1. 200499
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12p"
/clone="RP11-52513"
BASE COUNT 55194 a 44888 c 42583 g 57034 t 800 others
ORIGIN
Query Match 44.3%; Score 435.8; DB 2; Length 200499;
Best Local Similarity 97.4%; Pred. No. 1,4e-113;
Matches 443; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Query 1 atgtgtgacccagctgggtgtgacctggaagaatccccctagatgatgaactaac 60
|||||

```

Db 186895 ATGTGTAGCACCAGTGGGTGTGACTGTGAAGAAATCCCTAGATGATGATGACCTTAAC 186954
Oy 61 accatagaattcaaaatcccgactactacacagcaatactgtcttaagaagaccct 120
Db 186955 ACCATAGAAATCAAAATCCCTGCTACTACACACGACATGTTCTTAAGAGCACCTCT 187014
Oy 121 gctctcttcaccaaagctgtgtagacaagaattgtctccagaagggcctggggaat 180
Db 187015 GCTCTCTTCTCACCNAAGCGCTGTAGAACAAAGATTGTCCACAGAGGGCCTGGGGAT 187074
Oy 181 tttcagaacaatgtagcaatgtagacaaggtgtcatgtgctgtcagaataatcccaatcagt 240
Db 187075 TCTTGACGAATAGTATGATGACAGAGAGGTGTCAATGCTTCCAGAAATCCCAATCCACT 187134
Oy 241 gagaagggcataacctgtgcagaagaagaatgtctcttggaagaactttttgagatag 300
Db 187135 GAGAAAGCCATTAACCTGTGCAAGAAAAGTCTTGTGGAAGCAATCTTTGGAGTAGTG 187194

RESULT 13
LOCUS HSA329005 699 bp DNA linear PRI 01-OCT-2001
DEFINITION Homo sapiens genomic sequence surrounding NotI site, clone
ACCESSION AJ329005
VERSION AJ329005.1 GI:15873423
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 699)
Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F.,
Podewski, R.M., Matsushin, Y.G., Kvasha, S.M., Gyanchandani, A.,
Murevenko, O.V., Protodopov, A.I., Kashuba, V.I., Kisselev, L.L.,
Wasserman, M., Wahlstedt, C. and Zabarovsky, E.R.
Analysis of NotI flanking sequences: a new tool for gene discovery
and verification of the human genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 699)
Zabarovsky, E.R.
AUTHORS Direct Submission
TITLE Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
JOURNAL Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
source Location/Qualifiers
1. 699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NL1-FM8C"
BASE COUNT 178 a 164 c 171 g 185 t 1 others
ORIGIN
Query Match 25.8%; Score 253.8; DB 9; Length 699;
Best Local Similarity 99.2%; Pred. No. 1.4e-61;
Matches 255; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 673 agaaagctgaagaagaatgaagcttgatgagccactccagatgagctgctactct 732
Db 416 ACACACCTGAAGAAAGATGAGCTTTGATGGGCCACTTCCAGAGATGGCTGTCTACTCT 357

```

```

Oy 733 gtcttcagaccatcacagaccaggttcctaatgggtgtgagaccacaggggagaatcacag 792
Db 356 GTTTTCAGACACATCACAGACAGGCTCTTAATGGGTGTGAGACCCCAAGGAGATCAGAG 297
Oy 793 gtcaaaagctcagaggtttaaggtctgctgttaataagatcacagggcacaagctcacagct 852
Db 296 GTCAAGCTCAGGGGCTTTAAGGCTGCCCTTTAATFAGACGTGACGGCCAAAGTCAACAGCT 237
Oy 853 attgacaaccccccagatgaacaggggtcctggtgctttggcaccagaactgaagaagaac 912
Db 236 ATTGACAAACCAACCCGATGACAGAGGCTCTTGGCACCAGTAACCTGAAGAGAAC 177
Oy 913 tttcgcacatgatcca 929
Db 176 TTCTGCCATGATGATCA 160

RESULT 14
LOCUS AK000127 1731 bp mRNA linear PRI 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ20120 fis, clone COL05912.
ACCESSION AK000127
VERSION AK000127.1 GI:7020010
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL05912.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (sites)
Kawabata, A., Hiki, J., Kobatake, N., Inagaki, H., Ikema, Y.,
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
TITLE 2 (bases 1 to 1731)
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
TITLE Submitted (15-FEB-2000) to the DDBJ/EMBL/GenBank databases. Sumit
Sugano, Institute of Medical Science, University of Tokyo, Departm
of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction: 5' - & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source Location/Qualifiers
1. 1731
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL05912"
/clone_lib="COL"
/tissue_type="colon"
/note="Cloning vector pMR18SFL3"
BASE COUNT 506 a 404 c 342 g 479 t
ORIGIN
Query Match 21.9%; Score 215.6; DB 9; Length 1731;
Best Local Similarity 98.2%; Pred. No. 1.3e-50;
Matches 218; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 733 gtcttcagaccatcacagaccaggttcctaatgggtgtgagaccacaggggagaatcacag 792
Db 1 GTTTTCAGACACATCACAGACAGGCTCTTAATGGGTGTGAGACCCCAAGGAGATCAGAG 60
Oy 793 gtcaaaagctcagaggtttaaggtctgctgttaataagatcacagggcacaagctcacagct 852
Db 296 GTCAAGCTCAGGGGCTTTAAGGCTGCCCTTTAATFAGACGTGACGGCCAAAGTCAACAGCT 237

```



```

* 82089 82188: gap of unknown length
* 82189 84156: contig of 1968 bp in length
* 84157 84256: gap of unknown length
* 84257 86357: contig of 2101 bp in length
* 86358 86457: gap of unknown length
* 86458 87963: contig of 1506 bp in length
* 87964 88063: gap of unknown length
* 88064 89282: contig of 1219 bp in length
* 89283 91047: contig of 1665 bp in length
* 91048 91147: gap of unknown length
* 91148 93054: contig of 1907 bp in length
* 93055 93154: gap of unknown length
* 93155 95177: contig of 2023 bp in length
* 95178 95277: gap of unknown length
* 95278 96910: contig of 1633 bp in length
* 96911 97010: gap of unknown length
* 97011 98848: contig of 1838 bp in length
* 98849 98949: gap of unknown length
* 98949 100455: contig of 1507 bp in length
* 100456 100555: gap of unknown length
* 100556 102754: contig of 2199 bp in length
* 102755 102854: gap of unknown length
* 102855 104635: contig of 1781 bp in length
* 104636 104735: gap of unknown length
* 104736 106420: contig of 1685 bp in length
* 106421 106520: gap of unknown length
* 106521 107623: contig of 1242 bp in length
* 107623 107863: gap of unknown length
* 107863 109530: contig of 1668 bp in length
* 109531 109630: gap of unknown length
* 109631 111109: contig of 1479 bp in length
* 111110 111210: gap of unknown length
* 111210 113098: contig of 1889 bp in length
* 113099 113198: gap of unknown length
* 113199 114273: contig of 1075 bp in length
* 114274 114373: gap of unknown length
* 114374 116703: contig of 2330 bp in length
* 116704 116803: gap of unknown length
* 116804 117940: contig of 1137 bp in length
* 117941 118040: gap of unknown length
* 118041 120245: contig of 2205 bp in length
* 120246 120345: gap of unknown length
* 120346 121494: contig of 1149 bp in length
* 121495 121594: gap of unknown length
* 121595 123183: contig of 1589 bp in length
* 123184 123283: gap of unknown length
* 123284 124580: contig of 1297 bp in length
* 124581 124680: gap of unknown length
* 124681 126245: contig of 1565 bp in length
* 126246 126345: gap of unknown length
* 126346 127733: contig of 1388 bp in length
* 127734 127833: gap of unknown length
* 127834 129130: contig of 1297 bp in length
* 129131 129230: gap of unknown length
* 129231 130800: contig of 1570 bp in length
* 130801 130900: gap of unknown length
* 130901 132501: contig of 1601 bp in length
* 132502 132601: gap of unknown length
* 132602 134552: contig of 1951 bp in length
* 134553 134652: gap of unknown length
* 134654 135941: contig of 1289 bp in length
* 135942 136041: gap of unknown length
* 136042 137656: contig of 1615 bp in length
* 137657 137756: gap of unknown length
* 137757 138840: contig of 1084 bp in length
* 138841 138940: gap of unknown length

```

Query Match 17.9%; Score 176; DB 2; Length 149350;
 Best Local Similarity 71.5%; Pred. No. 4, 2e-39;
 Matches 289; Conservative 0; Mismatches 105; Indels 10; Gaps 4;

```

Db 96479 CCCCCGAGGGCGATGATGAGCGACCGACATAGAGTTCAAAAATCTGCGCTTTATGCCA 96420
QY 95 gacatcagtcctcaagagacacccgcctctctcaccacaaagctgtagaacaagaa 154
Db 96419 GACACCATGCTTCAAGAGACCCCGCGTGTCTTCCCAAGCTCTCCAGGACCCAGAA 96360
QY 155 gttgtccagagagggcctgaggaaattgttcagaaatgcatgtagagagtgcat 214
Db 96359 GTTGTGTCCAGAAACACATGSGGAGCTTGTGCT--CAACTGACTCTGACACAGATATCGT 96303
QY 215 ggccttcagaaattcccaatccagtcagtagaagggccataaacttggcagaagaaagtctt 274
Db 96302 TGCCCTGTAGAGATTCCTCTCCAGTGAAGGACATTAAGCTTGGGCAAGAAAGTCTT 96243
QY 275 cttagaagcattcttgagc---agtggagaggaagaattggcagaagccttgcca 331
Db 96242 CTTGAGAGAACTCTTCAAGTAAACAGAGAGAGAGAGAGCCGCGAGCTCCCAAAAGG 96183
QY 332 agtctcgtcagagtcagaaagacgttggaatccaagatcgcacagccagcagtggt 391
Db 96182 AGATCCATGCTCAAGGCGCCCTTCCCGGTAGAGGGGAGGAGGTAGG---AACCAGACTGCT 96126
QY 392 ccaggtgctcttcttaacgttgagcagtgcttggaagcatgaagct 435
Db 96125 CCAGATCCCTGTCCAGCGGTGAGAGAGAGTGCCTGAGAGAGTGAAGCT 96082

```

Search completed: June 20, 2002, 05:05:07
 Job time: 9107 sec

QY 36 cccctagatgatgactaaccataagaatc-aaaatcctgcgtactacaccca 94

DR WPI; 2001-488882/53.
DR P-PSDB; AAE07040.

DR P-PSDB; AAEO7040.

PT	Novel isolated human BCL-X-like polynucleotide, useful in therapeutic,
PT	diagnostic and pharmacogenic applications -

Claim 2a; Page 30; 33pp; English.

CC The present sequence is a cDNA encoding human BCL-X-like protein.
CC The BCL-X-like polynucleotides are useful in therapeutic, diagnostic
CC and pharmacogenetic applications. They are useful for screening drugs
CC effective in the treatment of symptomatic or phenotypic manifestations
CC perturbing the normal function of protein in the body and also for
CC treating physiological disorders and diseases. The BCL-X-like
CC polynucleotides are useful in conjunction with polymerase chain
CC reaction to screen libraries, isolate clones, to prepare cloning
CC and sequencing templates and as hybridisation probes for assessing
CC gene expression patterns.

50 Sequence 984 BP; 289 A; 230 C; 250 G; 215 T; 0 other;

every match	100.0%;	Score 984;	DB 22;	Length 984;
-------------	---------	------------	--------	-------------

Best Local Similarity 100.0%; Pred. No. 6,6e-299;

Matches	304;	conservative	0;	mismatches	0;	indels	0;	gaps	0;
---------	------	--------------	----	------------	----	--------	----	------	----

QY	1	atctgtgacacccagctgggtgttgacctggaagaatcccccatagtgtgttgacctaac	60
Db	1	atgtgtgacacccagctgggtgttgacctggaagaatcccccatagtgtgttgacctaac	60
QY	61	accatagaaattcaaaatcctcgcctactataccacagacatagtcttctaagagcaacct	120
Db	61	accatagaattcaaaatcctcgcctactataccacagacatagtcttctaagagcaacct	120
QY	121	gctctctccacccaagaagcgtctgtgagacaagaagttgtcccaagagggccctggggat	180
Db	121	gctctctccacccaagaagcgtctgtgagacaagaagttgtcccaagagggccctggggat	180
QY	181	tgttcagcaaaatgagttcatatgacagaggtgtcatatgacctgacgaatatcccaatccagt	240
Db	181	tgttcagcaaaatgagttcatatgacagaggtgtcatatgacctgacgaatatcccaatccagt	240
QY	241	gagaaagagccaataaacctctggccaagaaagaattctctctggagaagaatctctcttgagtagtg	300
Db	241	gagaaagagccaataaacctctggccaagaaagaattctctctggagaagaatctctcttgagtagtg	300
QY	301	gagaagaagaagatctgcagagcagcgtctgcacaagtgctctgtctcaaggttcaaaagagcgttg	360
Db	301	gagaagaagaagatctgcagagcagcgtctgcacaagtgctctgtctcaaggttcaaaagagcgttg	360
QY	361	gaataccaagaatctgcagacagccagccagttgttccaagtgctcttcttaacgtggagcgagctgc	420
Db	361	gaataccaagaatctgcagacagccagccagttgttccaagtgctcttcttaacgtggagcgagctgc	420
QY	421	tttgagacatgaagctgtgagaccccaagaatatttccaatctgcacacgagtagctgaataat	480
Db	421	tttgagacatgaagctgtgagaccccaagaatatttccaatctgcacacgagtagctgaataat	480
QY	481	gtttattctctgcgccaaccaaccaagaacgagccagggagaggtctcaagttccaagaagat	540
Db	481	gtttattctctgcgccaaccaaccaagaacgagccagggagaggtctcaagttccaagaagat	540
QY	541	tttgtaactagaggtctctctctccaggtccaaggtccaagctggacctgaagttaaatct	600
Db	541	tttgtaactagaggtctctctctccaggtccaaggtccaagctggacctgaagttaaatct	600
QY	601	aagaaagaatgaagaagaacaataactactgccaanaattgtttagctgtctgaataatccaaga	660
Db	601	aagaaagaatgaagaagaacaataactactgccaanaattgtttagctgtctgaataatccaaga	660
QY	661	gattcagtttgaaagaagaagctgaaagaagaataaagctcttgtagtgccaattccaagatgag	720
Db	661	gattcagtttgaaagaagaagctgaaagaagaataaagctcttgtagtgccaattccaagatgag	720

OY	721	ctgtccctactctgttttcaaacaccatcaaacacaggctctcctaagtggttggaacccagg	780
Db	721	ctgtccctactctgttttcaaacaccatcaaacacaggctctcctaagtggttggaacccagg	780
OY	781	ggaagaatcagaagctcaaaagctcaaggctcttaaaggctgcgccttghaatagacgtcaagcc	840
Db	781	ggaagaatcagaagctcaaaagctcaaggctcttaaaggctgcgccttghaatagacgtcaagcc	840
OY	841	aagctcacagatatttgaacaaccaccgcgatbaaacagggctcttgaggctttgacacaagtac	900
Db	841	aagctcacagatatttgaacaaccaccgcgatbaaacagggctcttgaggctttgacacaagtac	900
OY	901	ctgaagaagaaactcttcgcgcattgcatccagcaagcaagcttgatctggagaaaataacttggg	960
Db	901	ctgaagaagaaactcttcgcgcattgcatccagcaagcaagcttgatctggagaaaataacttggg	960
OY	961	atatcacatgaagaagatagaactga 984	
Db	961	atatcacatgaagaagatagaactga 984	

RESULT	2
AAH22582	
ID	AAH22582 standard; cDNA; 1179 BP.
vv	

AC AAH22582;

DT 07-SEP-2001 (first entry)

Human Bcl-1 polypeptide encoding cDNA.

KW Bcl-G; cancer; cancer therapy; oncogene; apoptosis; Bcl-GI; cytostatic; anti-invasive; 19-19; 3; human

XX **am** diuretoproloic; cefalosomine 12p12.3; human; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers

FT	CDS	196..1179
ET		/*+acc=

```

E1      /tag= a
E2      /product= "Bcl-G1"
E3      E2

```

XX

PN WO200144282-A2.

21-JUN-2001

13-DEC-2000: 2000WO-US33793.

XX

PR 14-DEC-1999; 99US-C

PA (BURN-) BURNHAM INS

Reed JC, Godzik A:

XX

DR P-PSDB; AAB85166.

[illegible]

PT PI novel polynucleotide encoding

XX Claim 3; Fig 1; 11pp; English.

QY	241	gagaagcgcaataaactcttgcaagaanaagctctcttggaagaactctttggagtagc	300
Dp	432	gagaaagcgccataaacctctggccaagaaaaagctctctctggaaagcattctcttggaaatagc	491
QY	301	gagaagggaaagatctcgcaagacagcgctcgccaaggtctctctgctcaaggtctcaagagcgctg	360
Dp	492	ggagaaaggaaagatcttcgcaagacagcgctcgccaaggtctctctgctcaaggtctcaagagcgctg	551
QY	361	gaatccaagatctcgcaacagcgcaagctggttcagagtgctcttctaacgctggagcagtgcc	420
Dp	552	gaatccaagatctcgcaacagcgcaagctggttcagagtgctcttctaacgctggagcagtgcc	611
QY	421	ttggagcatgaaagctgtggaccccaagaagctattccattgccaacggagtagctaaatt	480
Dp	612	ttggagcatgaaagctgtggaccccaagaagctattccattgccaacggagtagctaaatt	671
QY	481	gtttattcctgtgcacacccaacgaagcgaccccaagcagaagagcttccaagtccaagaagatt	540
Dp	672	gtttattcctgtgcacacccaacgaagcgaccccaagcagaagagcttccaagtccaagaagatt	731
QY	541	tttgaacatgaaaggtctctctctccaaactccaagaagcgaagcgctgtagcttcaagttct	600
Dp	732	tttgaacatgaaaggtctctctctccaaactccaagaagcgaagcgctgtagcttcaagttct	791
QY	601	aagaaagatgaaagaaagaaacaataactgccccaaatctgttgagctgtctgaaatatccaggaa	660
Dp	792	aagaaagatgaaagaaagaaacaataactgccccaaatctgttgagctgtctgaaatatccaggaa	851
QY	661	gatcagctgtggaaagaa-----	676
Dp	852	gatcagctgtggaaagaaagaaacactgcccctcatcccatctccctgtgttgacaaccagcattc	911
QY	677	-----	676
Dp	912	caaggttttccaacaggaatggttgatgagcctgcaattgagctaaagaattgaactctgtgc	971
QY	677	-----agctgaagaa	687
Dp	972	tgcctcgttgagaccaaagctactgtactgagtgctacttcttctgtacacaagctgaaagaa	1031
QY	688	gataaggtcttgaatggcgcaatcccaagatgagggctgtccatactctgtttccaagcattc	747
Dp	1032	gataaggtcttgaatggcgcaatcccaagatgagggctgtccatactctgtttccaagcattc	1091
QY	748	acagaacaaagcttctaataatgggtgtgtgaccccaaggggaaatcagaagtgccaagctcaaggtc	807
Dp	1092	acagaacaaagcttctaataatgggtgtgtgaccccaaggggaaatcagaagtgccaagctcaaggtc	1151
QY	808	tttaaggtcgccctgttaataagcgttcaagggccaaggtcaaaagctattgaaacaaccacgg	867
Dp	1152	tttaaggtcgccctgttaataagcgttcaagggccaaggtcaaaagctattgaaacaaccacgg	1211
QY	868	atgaaacaaaggtcccttggtgcttcttgccaccaagtaaccgtgaaagaaactcttcgcacatgac	927
Dp	1212	atgaaacaaaggtcccttggtgcttcttgccaccaagtaaccgtgaaagaaactcttcgcacatgac	1271
QY	928	cagcgacacaggtgagatggagaaaaaataactcttggatataccatgaaagaagtacgactgca	984
Dp	1272	cagcgacacaggtgagatggagaaaaaataactcttggatataccatgaaagaagtacgactgca	1328

RESULT	4
AA	AA590453
ID	AA590453 standard; cDNA; 1917 BP.
XX	
AC	AA590453; ~
XX	
DF	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #26257.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KN	food supplement; medical imaging; diagnostic; genetic disorder; ss

XX	Homo sapiens.
OS	
XX	W0200175067-A2.
PN	
XX	11-OCT-2001.
PD	
XX	
PF	30-MAR-2001; 2001WO-0508631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HISE-) HISEQ INC.
XX	
PI	Dmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG26266.
XX	
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PI	biodiversity -

Claim 1; SEQ ID No 26257; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. MAS64197-MA954564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at http://wipo.int/pub/published_pot_sequences.

50 Sequence 1917 BP; 553 A; 443 C; 439 G; 482 T; 0 other;

Query Match	80.4%	Score 791.6	DB 23	Length 1917
Best Local Similarity	87.7%	Pred. No. 3.3e-236		
Matches 975, Conservative	0	Mismatches 9	Indels 128	Gaps 5

Oy	1	atgtgtaacaccagttgggtg-tgacccggaagaatcccccataatgataatgataccctaa	59
Db	96	atgtgtaacaccagttggtgtgtgtgcacctggaagaaatcccccataatgataatgataccctaa	155
Oy	60	caccatagaattcaaatctctgcgctactataccagaacat--catgtcttcaagaagacc	117
Db	156	caccatagaattcaaaaatccctgcgctctctactaacagaacatcatgcttcttcaagaagacc	215
Oy	118	cctgtctc-cttctccaccaaagcgcgcgtgagaacaagaagttgttccagaaggggcctggg	176
Db	216	cctgtctgcgtcttccaccaaagcgcgcgtgagaacaagaagttgttccagaaggggcataagg	275
Oy	177	gaattgttccaaatagatgatgagaagaggtgtcctgtgcttgcgaagaattcccaatc	236
Db	276	gaattgttccagaatagtgatgagacagaggtgtcctgtgcttgcgaagaattcccaatc	335
Oy	237	caatgagaagggcatalaaccttggcaagaagaattcttcttggaaagcatcttcttggagt	296

D	b	336	caagtgaagaagccataaccttggcgaagaanaagctctcttcttggaagaacattctttagagt	395
O	y	297	agtcggagaagaagatctcgacagacacgccttcgcaaggtctctctcagggccaagaagac	356
D	b	396	agtggagaagaagaattcgcagagcagcgcctgcgaaggtctctctcagggccaagaagac	455
O	y	357	gttcgaataccaagatctgcacagcagcagctggctccaggtctctctcctaagctggagca	416
D	b	456	gttcgaataccaagatctgcacagcagcagctggctccaggtctctctcctaagctggagca	515
O	y	417	ggtcgttgagcagctgaa-----	432
D	b	516	gttccttgagcagctgaagctgctgtacaagaagcccaagagtgcccttagaggttgagacacc	575
O	y	433	-----	432
D	b	576	aacctctgtccaaaacgcacccttaactgtcttcctaagtcctcgtgtaattractaatgt	635
		433	-----gctgtgagccccaagctaatcttccatttcgccaacccagtagc	473
		636	taagcttccaatgttacaatgctgtgagccccaagctaatcttccatttcgccaacccagtagc	695
O	y	474	tgaataatgtttatctcttcgcccacacccaagaagcagcccaagcagaaggtcttcaagtcaca	533
D	b	696	tgaataatgtttatctcttcgcccacacccaagaagcagcccaagcagaaggtcttcaagtcaca	755
O	y	534	agaagattttctgaactcgaaggtctctccttcacagctcccaagcagcagctgcctgttagcttc	593
D	b	756	agaagattttctgaactcgaaggtctctccttcacagctcccaagcagcagctgcctgttagcttc	815
O	y	594	aagttcttaagaagaatgtgaagaagaacaataactag-ccaanaattgttgagctgtcgtgaat	652
D	b	816	aagttcttaagaagaagaatgtgaagaagaacaataactagccaanaattgttgagctgtcgtgaat	875
O	y	653	attcagcgaaatacgaattgggaagaagaagctggaagaagaataagctttgatggtgccacttc	712
D	b	876	attcagcgaaatacgaattgggaagaagaagctggaagaagaataagctttgatggtgccacttc	935
O	y	713	aggaatggctgtctcctactcctgtttccaagaccatcacagaccagagctcctaattgggtgctg	772
D	b	936	aggaatggctgtctcctactcctgtttccaagaccatcacagaccagagctcctaattgggtgctg	995
O	y	773	accocagggagaaatcagaagctgccaagcttcaaggtcttlaagctgtgccttgtaataagacg	832
D	b	996	accocagggagaaatcagaagctgccaagcttcaaggtcttlaagctgtgccttgtaataagacg	1055
		833	tcaagcgccaagctcacagctatttgacaacaaccocgatgaacaaggtctcctggcttggca	892
		1056	tcaagcgccaagctcacagctatttgacaacaaccocgatgaacaaggtctcctggcttggca	1115
O	y	893	ccaagtaacctgaaaggaactcttcgcataagatccagcagcagaaggttgatgggaaaaaa	952
D	b	1116	ccaagtaacctgaaaggaactcttcgcataagatccagcagcagaaggttgatgggaaaaaa	1175
O	y	953	tacttgatatcacatgaagaagttagactga 984	
D	b	1176	tacttgatatcacatgaagaagttagactga 1207	
RESULT 5				
AADI3236				
ID	AADI3236 standard; cDNA; 759 BP.			
XX	AADI3236;			
XX	23-OCT-2001 (first entry)			
XX	Human BCL-X-like protein encoding cDNA #2.			
DE	Human; BCL-X-like protein; therapy; physiological disorder; ss.			
XX	Homo sapiens.			
XX				

FH	Key	Location/Qualifiers
FT	CDS	1..759
FT		/tag= a
FT		/product= "Human BCL-X-like protein #2"
XX		
PM	WO200157213-A2.	
XX		
PD	09-AUG-2001.	
XX		
PF	31-JAN-2001; 2001MO-US03446.	
XX		
PR	04-FEB-2000; 2000US-0180412.	
XX		
PA	(LEXI-) LEXICON GENETICS INC.	
PI	Donoho G, Hilburn E, Turner CA, Friedrich G, Abuin A, Zambrowicz B;	
PI	Sands AT;	
XX		
DR	WPI; 2001-488862/53.	
DR	P-PSDB; AAEO7041.	
XX		
PT	Novel isolated human BCL-X-like polynucleotide, useful in therapeutic,	
XX	diagnostic and pharmacogenic applications -	
PS	Disclosure; Page 31; 33pp; English.	
CC	The present sequence is a cDNA encoding human BCL-X-like protein.	
CC	The BCL-X-like polynucleotides are useful in therapeutic, diagnostic	
CC	and pharmacogenic applications. They are useful for screening drugs	
CC	effective in the treatment of symptomatic or phenotypic manifestations	
CC	perturbing the normal function of protein in the body and also for	
CC	treating physiological disorders and diseases. The BCL-X-like	
CC	polynucleotides are useful in conjunction with polymerase chain	
CC	reaction to screen libraries, isolate clones, to prepare cloning	
CC	and sequencing templates and as hybridisation probes for assessing	
CC	gene expression patterns.	
SQ	Sequence 759 BP; 216 A; 182 C; 185 G; 176 T; 0 other;	
<hr/>		
Query Match 68.9%; Score 678; DB 22; Length 759;		
Best Local Similarity 100.0%; Pred. No. 1.1e-202;		
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1 atgttgagaccacgctgggtgtgaactggagaagaatccccctagtatgatgacctaac 60	
DB	1 atgttgagaccacgctgggtgtgaactggagaagaatccccctagtatgatgacctaac 60	
QY	61 accatagaattccaatacctctgcctctacacacagacatcatcttccaagagaccct 120	
DB	61 accatagaattccaatacctctgcctctacacacagacatcatcttccaagagaccct 120	
QY	121 gctctcttcaccaaaagctgtcgagaacaagaattttgccagaaggcgcttggggaat 180	
DB	121 gctctcttcaccaaaagctgtcgagaacaagaattttgccagaaggcgcttggggaat 180	
QY	121 gctctcttcaccaaaagctgtcgagaacaagaattttgccagaaggcgcttggggaat 180	
DB	121 gctctcttcaccaaaagctgtcgagaacaagaattttgccagaaggcgcttggggaat 180	
QY	181 tgttcagcaaatgagtcatagtgacagaggtgtcatagtgcttcgcagaatcccaatcagt 240	
DB	181 tgttcagcaaatgagtcatagtgacagaggtgtcatagtgcttcgcagaatcccaatcagt 240	
QY	241 gagaaagccataaacctctggcagaagaagaattctctctgaaaagcatctcttgagtagtg 300	
DB	241 gagaaagccataaacctctggcagaagaagaattctctctgaaaagcatctcttgagtagtg 300	
QY	301 gagaagaagaattctgcagagcacgcttccaaagttctctgtcgaaggtccaaggaagcttg 360	
DB	301 gagaagaagaattctgcagagcacgcttccaaagttctctgtcgaaggtccaaggaagcttg 360	
QY	361 gaataccaagaattctgcagagcacgcttccaaagttctctgtcgaaggtccaaggaagcttg 420	
DB	361 gaataccaagaattctgcagagcacgcttccaaagttctctgtcgaaggtccaaggaagcttg 420	
QY	421 tttagagcattaagaattctgagaccccaagaattcatttcattgacgcaaacgagtagctgaaatt 480	
DB	421 tttagagcattaagaattctgagaccccaagaattcatttcattgacgcaaacgagtagctgaaatt 480	

Db 421 ttgagcagtcgaagctgtgaggaccacaagtcattcttcattgcaccacggagtagtgtaatt 480
 |||
 Oy 481 gttattctccggcacacacacacagaaccagcgaggaagcttcaagttcaagaagatt 540
 |||
 Db 481 gttattctccggcacacacacacagaaccagcgaggaagcttcaagttcaagaagatt 540
 |||
 Oy 541 ttgttaactgagggctctctctcttcacgtcccaagcgacgtgctgtagcttcaattct 600
 |||
 Db 541 ttgttaactgagggctctctctcttcacgtcccaagcgacgtgctgtagcttcaattct 600
 |||
 Oy 601 aagaagaatgaagaagaacaatactacgccaatatgttagctgtcgaataattgaaga 660
 |||
 Db 601 aagaagaatgaagaagaacaatactacgccaatatgttagctgtcgaataatttcagga 660
 |||
 Oy 661 gatcagtttggaagaagaag 678
 |||
 Db 661 gatcagtttggaagaagaag 678
 |||

PN WO200144282-A2.
 XX
 PD 21-JUN-2001.
 XX
 PF 13-DEC-2000; 2000WO-US33793.
 XX
 PR 14-DEC-1999; 99US-0461641.
 XX
 PA (BURN-) BURNHAM INST.
 XX
 Reed JC, Godzik A;
 XX
 WPI; 2001-398125/42.
 DR P-PSDB; AAB85167.
 XX
 PT Novel polynucleotide encoding Bcl-6 polypeptide, useful for modulating
 PT apoptosis, and for diagnosing and treating cancer -
 XX
 SS Claim 3; Fig 3; 117pp; English.

Sequence 954 BP; 267 A; 237 C; 240 G; 210 T; 0 other;

Query Match	68.7%	Score	676.4	DB	22	Length	954
Best Local Similarity	99.9%	Pred.	No. 3.9e-202				
Matches	677	Conservative	1			Indels	0
						Gaps	0

QY	1	atggttaacacacagttggtgtgtgaccttggaaagaaatcccccatagtgatgtgaccttaac	60
Dp	196	atgtgtgacacacagttggtgtgtgaccttggaaagaaatcccccatagtgatgtgaccttaac	255
QY	61	accatagaatcataaattcctgcgtactataccacagacatactgtcttcaagaagacacct	120
Dp	256	accatagaatcataaattcctgcgtactataccacagacatactgtcttcaagaagacacct	315
QY	121	gctctcttctcaacaaagctgtctgaagaaagaagattgttccagaaaggccttgggaat	180
Dp	316	gctctcttctcaacaaagctgtctgaagaaagaagattgttccagaaaggccttgggaat	375
QY	181	tgttcacgaacaaatgtcatgtgacagagagttgtcaatgctcttgcagaaattcccaatcagt	240
Dp	376	tgttcacgaacaaatgtcatgtgacagagagttgtcaatgctcttgcagaaattcccaatcagt	435
QY	241	gagaaaggccataaacctctgtgaaagaaaaaagtctctcttggaaagcatctcttggatgtg	300
Dp	436	gagaaaggccataaacctctgtgaaagaaaaaagtctctcttggaaagcatctcttggatgtg	495
QY	301	gagaagaagaaagatttcagagagacagcctctgcagaaagttctctcagaggltcaagaagactg	360
Dp	496	gagaagaagaaagatttcagagagacagcctctgcagaaagttctctcagaggltcaagaagactg	555
QY	361	gaataaccaagatttcgcacagccagacatgtgtccagtggtcttcttaacgttggagcagttgc	420
Dp	556	gaataaccaagatttcgcacagccagacatgtgtccagtggtcttcttaacgttggagcagttgc	615
QY	421	tttgagacatgtagctgtctggagcccaagaagtcattctcattgcacacagagtagctgaaatt	480
Dp	616	tttgagacatgtagctgtctggagcccaagaagtcattctcattgcacacagagtagctgaaatt	675
QY	481	gtttattctctgtgccacacacacaaagcgaaacccaagcagaagagcttcaagttccaaagagatt	540
Dp	676	gtttattctctgtgccacacacacaaagcgaaacccaagcagaagagcttcaagttccaaagagatt	735
QY	541	tttgttaactgagggctctctcttccttcagctccaaaggcagctgtccctgttagcttcaagttct	600
Dp	736	tttgttaactgagggctctctcttccttcagctccaaaggcagctgtccctgttagcttcaagttct	755
QY	601	aagaaagaatgaaagaagaacaatactagtcacaaatttgtttagctgtcgtgaataattcagga	660
Dp	796	aagaaagaatgaaagaagaacaatactagtcacaaatttgttagctgtcgtgaataattcagga	855
QY	661	gattcagtttggaaaaagaag 678	
Dp	856	gattcagtttggaaaaagaag 873	

RESULT 7

AAH22670
ID AAH22670 standard; DNA; 987 BP

AAH22670;

DT 07-SEP-2001 (first entry)

Mouse Bcl-6 polypeptide encoding cDNA.

Bcl-G; cancer; cancer therapy; oncogene; apoptosis; cytostatic;

KW antiapoptotic; chromosome 12p12.3; mouse; ss.

Mus sp.

FH	Key	Location/Qualifiers
FT	CDS	1..987

PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220863.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228282.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236870.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239835.
PR 13-OCT-2000; 2000US-0239837.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI: 2001-465570/50.
XX P-FSDB; AAM95378.
XX
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX
XX
XX Claim 1; SEQ ID NO 1349; 1297bp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
XX invention.
XX
SQ Sequence 632 BP; 190 A; 143 C; 166 G; 128 T; 5 other;

XX 01-FEB-2002 (first entry)
 XX Human foetal liver single exon nucleic acid probe #13287.
 DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 XX Homo sapiens.
 OS WO200157277-A2.
 XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00669.
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver -

Claim 4; SEQ ID NO 13287; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human foetal liver. The present sequence is a single exon nucleic acid probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;

Query Match 39.4%; Score 388; DB 22; Length 388;

Best Local Similarity 100.0%; Pred. No. 1.3e-111; Mismatches 0; Indels 0; Gaps 0;

46 gatgatgacctaacaacacatagaatcaaatcctgcctactacacagacatcatgctc
 1 gatgatgacctaacaacacatagaatcaaatcctgcctactacacagacatcatgctc
 106 ttcgaagacacccctctctctctcaccaaagtctgtgagaacaagaattgtccag
 61 ttcgaagacacccctctctctctcaccaaagtctgtgagaacaagaattgtccag
 166 aggggctctgggagatgtttagcaaatgagtcagagaggggtgtatgctgtgcaaa
 121 aggggctctgggagatgtttagcaaatgagtcagagaggggtgtatgctgtgcaaa
 226 aattcccaatccagtgtagaagggcacaataccttggcaagaataatctcttgaaga
 181 aattcccaatccagtgtagaagggcacaataccttggcaagaataatctcttgaaga
 286 ttccttggagtagtgtagaaggaagatctgcagagcagcctgtgccaaagtctctcag
 241 ttccttggagtagtgtagaaggaagatctgcagagcagcctgtgccaaagtctctcag
 346 ggtcaagaagacgttggataatccaagattcgacagcagcagtggttccaggtgtcttct

Db 301 ggtcaagaagacgttggataatccaagattcgacagcagcagtggttcttct 360
 Oy 406 aacgtgagcagtgcttggagacatgaag 433
 Db 361 aacgtgagcagtgcttggagacatgaag 388

RESULT 11

ABA32090
 ID ABA32090 standard; DNA; 388 BP.

AC ABA32090;

DT 23-JAN-2002 (first entry)

DE Probe #10556 for gene expression analysis in human heart cell sample.

KW Human; gene expression; heart; microarray; vascular system; probe;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease; ss.

OS Homo sapiens.

WO200157274-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00666.

04-FEB-2000; 2000US-0180312.
 26-MAY-2000; 2000US-0207456.
 30-JUN-2000; 2000US-0608408.
 03-AUG-2000; 2000US-0632366.
 21-SEP-2000; 2000US-0234687.
 27-SEP-2000; 2000US-0236359.
 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI: 2001-488899/53.

Single exon nucleic acid probes for analyzing gene expression in human hearts -

Claim 4; SEQ ID NO 10556; 530pp; English.

The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosticating diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;

Query Match 39.4%; Score 388; DB 22; Length 388;

Best Local Similarity 100.0%; Pred. No. 1.3e-111; Mismatches 0; Indels 0; Gaps 0;

46 gatgatgacctaacaacacatagaatcaaatcctgcctactacacagacatcatgctc
 1 gatgatgacctaacaacacatagaatcaaatcctgcctactacacagacatcatgctc
 106 ttcgaagacacccctctctctctcaccaaagtctgtgagaacaagaattgtccag 165

Db 61 ttcaagagcaccctgctctctctctcaccacaagctgctgagacaagaattgtccag 120
 Qy 166 aggggctgggggagattgttgaagcaatgacatgacagaggtgtcgtgttcaga 225
 Db 121 aggggctgggggagattgttgaagcaatgacatgacagaggtgtcgtgttcaga 180
 Qy 226 aattcccaatccagctgagagagccataaacttgcgaagaaatctcttgcgaaga 285
 Db 181 aattcccaatccagctgagagagccataaacttgcgaagaaatctcttgcgaaga 240
 Qy 286 ttcttggagtagtggagaaagatctgcagagacgctgccaaggtctctgtcag 345
 Db 241 ttcttggagtagtggagaaagatctgcagagacgctgccaaggtctctgtcag 300
 Qy 346 ggtcaaaagacgttggaaataaccagatctgcacagccagctgttccaggtcttct 405
 Db 301 ggtcaaaagacgttggaaataaccagatctgcacagccagctgttccaggtcttct 360
 Qy 406 aacgtgagacagtgcttggagcatgaag 433
 Db 361 aacgtgagacagtgcttggagcatgaag 388

RESULT 12
 AAK13408
 ID AAK13408 standard; DNA; 388 BP.

AC AAK13408;
 DT 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe SEQ ID NO: 13399.
 XX
 KM Human; brain expressed exon; gene expression analysis; probe;
 KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KM epilepsy; cancer; ss.
 OS Homo sapiens.
 XX
 PN WO200157275-A2.

09-AUG-2001.
 30-JAN-2001; 2001WO-US00667.
 04-FEB-2000; 2000US-0180312.
 26-MAY-2000; 2000US-0207456.
 30-JUN-2000; 2000US-0608408.
 03-AUG-2000; 2000US-0632366.
 21-SEP-2000; 2000US-0234687.
 27-SEP-2000; 2000US-0236359.
 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-483446/52.

Single exon nucleic acid probes for analyzing gene expression in human brains -

Example 4; SEQ ID NO: 13399; 650pp + Sequence Listing; English.

The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.

Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;

Query Match 39.4%; Score 388; DB 22; Length 388;
 Best Local Similarity 100.0%; Pred. No. 1.3e-111;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 46 gatgatgacctaaacacatagaaatcacaatctctgctactacacagacatcatgtc 105
 Db 1 gatgatgacctaaacacatagaaatcacaatctctgctactacacagacatcatgtc 60
 Qy 106 ttcaagagacccctgctctctcttccaccaagctgctggaacaagaattgtccag 165
 Db 61 ttcaagagacccctgctctctcttccaccaagctgctggaacaagaattgtccag 120
 Qy 166 aggggctgggggagattgttcaagaatgagtcatagcaagaaggtgtatggtccttcaga 225
 Db 121 aggggctgggggagattgttcaagaatgagtcatagcaagaaggtgtatggtccttcaga 180
 Qy 226 aattcccaatccagctgagagagccataaacttgcgaagaaatctcttgcgaaga 285
 Db 181 aattcccaatccagctgagagagccataaacttgcgaagaaatctcttgcgaaga 240
 Qy 286 ttcttggagtagtggagaaagatctgcagagacgctgccaaggtctctgtcag 345
 Db 241 ttcttggagtagtggagaaagatctgcagagacgctgccaaggtctctgtcag 300
 Qy 346 ggtcaaaagacgttggaaataaccagatctgcacagccagctgttccaggtcttct 405
 Db 301 ggtcaaaagacgttggaaataaccagatctgcacagccagctgttccaggtcttct 360
 Qy 406 aacgtgagacagtgcttggagcatgaag 433
 Db 361 aacgtgagacagtgcttggagcatgaag 388

RESULT 13
 AAK39145
 ID AAK39145 standard; DNA; 388 BP.

AC AAK39145;
 DT 06-NOV-2001 (first entry)
 DE Human bone marrow expressed single exon probe SEQ ID NO: 13702.

XX
 KM Human; bone marrow expressed exon; gene expression analysis; probe;
 KM microarray; cancer; leukemia; lymphoma; myeloma; ss.
 OS Homo sapiens.
 XX
 PN WO200157276-A2.

09-AUG-2001.
 30-JAN-2001; 2001WO-US00668.

04-FEB-2000; 2000US-0180312.
 26-MAY-2000; 2000US-0207456.
 30-JUN-2000; 2000US-0608408.
 03-AUG-2000; 2000US-0632366.
 21-SEP-2000; 2000US-0234687.
 27-SEP-2000; 2000US-0236359.
 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488900/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow -

```

XX Example 4; SEQ ID NO: 13702; 658bp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SO Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;

Query Match          39.4%; Score 388; DB 22; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.3e-111;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gatgtgacctaacacccatagaattcaaatcctgcctctactacacagatcatgtc 105
DB 1 gatgatgacctaacacccatagaattcaaatcctgcctctactacacagatcatgtc 60
    106 ttcaagaagacccctctctctctcaccacaaagctgctgagaaagaagttgtccag 165
    61 ttcaagaagacccctctctctctcaccacaaagctgctgagaaagaagttgtccag 120
QY 166 aggggcttgagggaattgttcagaatgatgatgacagaagtgatgacctgacaga 225
DB 121 aggggcttgagggaattgttcagaatgatgatgacagaagtgatgacctgacaga 180
QY 226 aattcccaatcagtgatgagaagggccataaactctggcagaagaagttctcttgaaagca 285
DB 181 aattcccaatcagtgatgagaagggccataaactctggcagaagaagttctcttgaaagca 240
QY 286 ttctttgagtagtgagagaagaagatttcgacagacagcctgcacaaagtctctgtcag 345
DB 241 ttctttgagtagtgagagaagaagatttcgacagacagcctgcacaaagtctctgtcag 300
QY 346 ggtcaaaagacgttggaatacaagaattcgacacagcagcagtggtccagtgcttct 405
DB 301 ggtcaaaagacgttggaatacaagaattcgacacagcagcagtggtccagtgcttct 360
QY 406 aacgttgagcagtgcttgagcatgaag 433
DB 361 aacgttgagcagtgcttgagcatgaag 388

RESULT 14
AA119954
ID AA119954 standard; DNA; 388 BP.
XX
XX AA119954;
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #9887 for gene expression analysis in human cervical cell sample.
DE
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00670.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.

```

```

PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 25; SEQ ID No 9887; 487bp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human Hela cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging
XX of diseases of the cervix, notably cervical cancer.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;

Query Match          39.4%; Score 388; DB 22; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.3e-111;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gatgatgacctaacacacatagaattcaaatcctgcctactacacagatcatgtc 105
DB 1 gatgatgacctaacacacatagaattcaaatcctgcctactacacagatcatgtc 60
    106 ttcaagaagacccctctctctctcaccacaaagctgctgagaaagaagttgtccag 165
    61 ttcaagaagacccctctctctctcaccacaaagctgctgagaaagaagttgtccag 120
QY 166 aggggcttgagggaattgttcagaatgatgatgacagaagtgatgacctgacaga 225
DB 121 aggggcttgagggaattgttcagaatgatgatgacagaagtgatgacctgacaga 180
QY 226 aattcccaatcagtgatgagaagggccataaactctggcagaagaagttctcttgaaagca 285
DB 181 aattcccaatcagtgatgagaagggccataaactctggcagaagaagttctcttgaaagca 240
QY 286 ttctttgagtagtgagagaagaagatttcgacagacagcctgcacaaagtctctgtcag 345
DB 241 ttctttgagtagtgagagaagaagatttcgacagacagcctgcacaaagtctctgtcag 300
QY 346 ggtcaaaagacgttggaatacaagaattcgacacagcagcagtggtccagtgcttct 405
DB 301 ggtcaaaagacgttggaatacaagaattcgacacagcagcagtggtccagtgcttct 360
QY 406 aacgttgagcagtgcttgagcatgaag 433
DB 361 aacgttgagcagtgcttgagcatgaag 388

RESULT 15
AA145152
ID AA145152 standard; DNA; 388 BP.
XX
XX AA145152;
XX
XX 17-OCT-2001 (first entry)
XX
XX Probe #13838 used to measure gene expression in human placenta sample.
DE
XX
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX
XX

```

OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 13838; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SNP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other:

Query Match 39.4%; Score 388; DB 22; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.3e-111;
Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gatgatgacctaacaccataagaatcctaactcctgctactacccaagacatcatgtc 105
Db 1 gatgatgacctaacaccataagaatcctaactcctgctactacccaagacatcatgtc 60
QY 106 ttcaagagacccccctgctctctctcaccacaagctgtcgaacaagaattgtccag 165
Db 61 ttcaagagacccccctgctctctctcaccacaagctgtcgaacaagaattgtccag 120
QY 166 aggggacctgaggaaattgttcagcaaatgagtcagacagaggtgtcatgacctgcaga 225
Db 121 aggggacctgaggaaattgttcagcaaatgagtcagacagaggtgtcatgacctgcaga 180
QY 226 aattcccaatccagtgagaaagggccataaccttggcaagaagaattcttctgaaagca 285
Db 181 aattcccaatccagtgagaaagggccataaccttggcaagaagaattcttctgaaagca 240
QY 286 ttcttggagtaattggaagaagaatttcgacagacagcctgcgaaggtctctgtcag 345
Db 241 ttcttggagtaattggaagaagaatttcgacagacagcctgcgaaggtctctgtcag 300
QY 346 ggtcaaaagacgttgcgaataccaagattcgacagccagcagtggtccagtgcttct 405
Db 301 ggtcaaaagacgttgcgaataccaagattcgacagccagcagtggtccagtgcttct 360
QY 406 aacgtgagagcagtgctgagcagatgaag 433
Db 361 aacgtgagagcagtgctgagcagatgaag 388

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 02:32:30 : Search time 2402.36 Seconds
(Without alignments)
5528.315 Million cell updates/sec

Title: US-09-771-961-1

Perfect score: 984

Sequence: 1 atcgtacgaccacagtggtg.....cacatgaagaagtagactga 984

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

al number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_iny:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	694.4	70.6	980	10	BI833102
2	603.4	61.3	763	10	BG718587
3	549	55.8	774	10	BG720221
4	527	53.6	1535	11	AK016997
5	527	53.6	1541	11	AK008682
6	527	53.6	1578	11	AK018579
7	527	53.6	2232	11	AK016670
8	506.6	51.5	715	10	BI838169
9	497.2	50.5	672	10	BG717835
10	418.4	42.5	703	10	BI908814
11	376	38.2	426	9	AM000827
12	359	36.5	676	10	BI652080
13	356.8	36.3	422	12	AO131601
14	342.6	34.8	984	9	BB013694
15	321.4	32.7	406	9	AA399486
16	303	30.8	451	10	BM433532
17	276.2	28.1	776	10	BI463753

18	255.8	26.0	394	10	BE631990
19	246.6	25.1	446	9	BB746476
20	246.6	25.1	920	10	BI453888
21	246.4	25.0	652	9	AA536718
22	243.6	24.8	680	10	BG245730
23	242.6	24.7	448	10	BF198484
24	234.4	23.8	697	10	BF584699
25	227.2	23.1	924	10	BG176106
26	209	21.2	666	9	BB628327
27	204.4	20.8	644	10	BM211865
28	203.8	20.7	1061	10	BF139422
29	201.2	20.4	403	10	BI342342
30	197	20.0	567	9	AM631581
31	190.6	19.4	922	9	BB615307
32	186	18.9	465	10	BF149673
33	149	15.1	423	10	BG895661
34	141.8	14.4	559	10	BM216619
35	133.6	13.6	437	9	AA061006
36	132	13.4	848	10	BI158784
37	130.6	13.3	397	9	AI614194
38	123	12.5	354	9	AM414655
39	109.2	11.1	376	9	AM484860
40	104.8	10.7	370	9	AM428115
41	97.6	9.9	216	10	BI360692
42	93.8	9.5	332	10	T69681
43	88.6	9.0	473	9	AA400686
44	87.4	8.9	256	9	AI606692
45	83	8.4	467	9	AM001213

ALIGNMENTS

RESULT 1
BI833102
LOCUS
DEFINITION
603090814F1 NIH_MGC_120 Homo sapiens CDNA clone IMAGE:5229819 5',
mRNA sequence.
BI833102
ACCESSION
BI833102.1 GI:15944652
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM11578 row: k column: 04
High quality sequence stop: 803.

FEATURES

source
1..980
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5229819"
/clone_lib="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-Sport6; Site: 1: NotI; Site: 2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is

Db	565	TTGGAGCATGAAGCTGTGGACCCCAAAATTCATTTCGACTTGCACACCAGACTACTGAANT	624
OY	481	gttattcttcggccaaccacaagaagcgaccccaaggcaggatggtcctaagtccaagaagatt	540
Db	625	GTTTATCTCTGGCCACCGACCAACGACGCCAGCAGGAGGCTTCAGTGCCAAAGAGATT	684
OY	541	tttgaactgagggctctccttcctccagctccaagcgagcgcttgtagttcaagttct	600
Db	685	TTTGTAAGCTGAGGGCTCTCTCTCCATCCAGCTCCAAGGCCACAGTGCCTGTAGCTTCAAGTCT	744
OY	601	aaga 605 	
Db	745	AAGAA 749	
RESULT	3		
BG720221			
LOCUS		774 bp	mRNA linear EST 08-MAY-2001
DEFINITION		M6269227.F1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4824760 5' ,	
VERSION		BG720221	
SOURCE		BG720221.1 GI:13999408	
KEYWORDS		EST.	
ORGANISM		human.	
REFERENCE		Homo sapiens	
AUTHORS		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
TITLE		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
JOURNAL		1 (bases 1 to 774)	
COMMENT		NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapds@email.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov Plate: LLM10736 row: m column: 17 High quality sequence stop: 769. Location/Qualifiers 1..774 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4824760" /clone_id="NIH_MGC_97" /lab_host="DH10B" /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcagag); Oligo-gt primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to Rof 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."	
BASE COUNT		225 a 185 c 198 g 166 t	
ORIGIN			
Query Match		55.8%; Score 549; DB 10; Length 774;	
Best Local Similarity		98.1%; Pred. No. 1.1e-143;	
Matches 609; Conservative 0; Mismatches 5; Indels 7; Gaps 5;			
OY	1	atgtgtagcaccaagtgtgtgatccctggaagaatacccctcatgatgatgaactaac 60	
Db	159	ATGTGTACACACAAGTGGGTGTGACCTGGAAGAATCCCTCATGATGATGATGATTAAC 218	
OY	61	acatatagaattcaaaatcctgcacctactacacagacatcatgttcttaagaacacct 120	

Db	Accession	Gene	Species	Strain	Library	Enrichment	Cloning	Reference
Db	219	ACCTAGATTTCAAAATTCCTGGCCCTACTACACAGACATCATCTCTTCAAGAGCACCCCT	278					
QY	121	gctctctctcccaaaagctctgtgaaacaagaagttgtgccagaagggccttggggat	180					
Db	279	GCCTCTTTCACCAAAAGCTGCTGAGAACAAAGATTGTCACAGAGGGGCGCTGGGAAAT	338					
QY	181	tgctcagcaaatgagtcagtgcagagaggtgtcattgctcttcagaaatctccaatcaagt	240					
Db	339	TGTTTCAGCAAAATGATGATGACAGAGGGTGTATGGCTCTTGCGAATAATCCCAATTCAGT	398					
QY	241	gaagaagcccaaaacatttggcaagaagaagctctcttggaaagcattctttggagtgatg	300					
Db	399	GAGAAAGCCCTAAACCTTGGCAAGAAAAGTCTCTTGGAAAGCATTTCTTGGAGTGGTG	458					
QY	301	gaagaaggaagatctgcagagcaagccttccaaagctctcgtcagaggtccaaagagcttg	360					
Db	459	GAGAAAGCAATGTCAGAGAGCAGCGCTCCCAAGTCTGCTCAGGTCGTAAGAGCAGTGG	518					
QY	361	gaataccaagaattcgcacagcagcagcagtggtccaggtgtcttcttaacgttggagcagtg	420					
Db	519	GAATACCAAAATTCGACAGCAGCAGAGGTGTCAGAGTGTCTTTCAACGTGAGACAGTGC	578					
QY	421	-tttggagcagtagaagctgttggagcccaagtgatcttccatttgcacacagtagtgaat	479					
Db	579	TTTGTGACATGAACTGTGGACCCCAAAAGTCAATTTCCATTGTCACACCGAGTAGCTGAAT	638					
QY	480	tgcttctctctgcgcaccacacacacagcagcagcagcagcagcagcagcagcagcagcagc	539					
Db	639	TGTTTACTCTGTC - ACACCAAGGCGA - CCAAGCAGAGGCTTCAAGTCCAAAGAGAT	695					
QY	540	tttttgtaactgaggggtctctctctcctcagctccaa - ggcacagtgctgtagcttcaagtt	598					
Db	696	TTTGTGACATGAGGGGTCTCTCTTCCAGCTCCAAAGGCGACAGTGTGTAG - TTTAGTT	753					
QY	599	ctagaagaagatgaagaagaac	619					
Db	754	CTAAGACATGAGAGAGAAC	774					
RESULT	4							
LOCUS	AK016997	1535 bp	mRNA	1 linear	HTC 19-JAN-2002			
DEFINITION	AK016997	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933430L01:BCL2-1like apoptosis inhibitor containing protein, full insert sequence.						
VERSION	AK016997	AK016997.1	GI:12856033					
KEYWORDS	HTC; CAP trapper.							
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library							
ORGANISM	Mus musculus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.							
AUTHORS	1 (sites)							
TITLE	Carninci, P. and Hayashizaki, Y.							
JOURNAL	High-efficiency full-length cDNA cloning							
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)							
PUBMED	99279253							
REFERENCE	10349636							
AUTHORS	2 (sites)							
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.							
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes							
MEDLINE	Genome Res. 10 (10), 1617-1630 (2000)							
PUBMED	20499374							
REFERENCE	11042159							
AUTHORS	3 (sites)							
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, O., Nishikawa, K., Katsunari, T., Tashiro, H., Itoh, M., Sumi, T., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,							

OY	301	gaga-...aggagaatttcgcagagcagcgctgcgaaggtcttcgtcccaaggtc-----	349
Db	471	GAGAAAGGAGGAAGGCGCTCCGAGCTCCCCAAAGAGATCCAGCTCAGGTCCTCAGAGGC	530
OY	350	---aaagcagcttgaataccaagaatttcgcagcagccacagcagtgctgcaggtcttct	405
Db	531	CCCTTCCTCCGGTAAAGACGGCGACAGTGGCTTCCTCCCAACACAGACTGGCCAGGCTCTCTGAGC	590
OY	406	aacttgagcagtgctgttgagcatbaaactcttgagccccaagatcatctccatcccaac	465
Db	591	AGTGTGAGACAGCCGCTCTGAGAGTAAGTTGGATTCCAAAGTGGCTTGTATATGCGAAC	650
OY	466	cgagtaagctgaatatgttattctctcgtgcacacacacacacagcgacgagcaggaagcttc	525
Db	651	AGAGGCGCTGAAATGTGTACTCTCGGCGCACACACAGATGTATCTCACAGCCAGGAGAGA	710
OY	526	aagtcacaagaagatllttgttaactbaaggtctctctcttcagcttcacaagcgagtgct	585
Db	711	AGCAAGCTCAAAAGAGAGGGGCTCTGGAGATTTCTGTATCTTCAAGTTTGAAGGAC-...-CTT	765
OY	586	gtagcttcaagttctcaagaagatlaaagaagacaaatctactagccaaaattbttagctg	645
Db	766	GTGACTCTAAG-...AATAAAGATGGAGAGAGCAAAATATATAGCAAGATGTGTAGCTG	821
OY	646	ctgaaatctcaagagatcagtttggaaagaaagctgaaagaaagtaagctttagtgagc	705
Db	822	CTGAATTTCTCGGGGGATCAGTTGGGAAGAGAGATTAAGAAAGAACAGGCTTTGATGAGAC	881
OY	706	caactccaagatagtgctgtctctactcgttcttcaagaacatcaacagcagtgctaatg	765
Db	882	AGCTTCAGAGAGGGGCTCTCTACTCAACGTTCAAGACCATACAGACCTGTGTCTCGAGG	941
OY	766	ggtgtgagaccccaaggaggaatcagaagttcaaagctcagggtcttaaggctgccttga	825
Db	942	GACGTGGACACCCAGAGGGAATTCAGAGGTCMAAGCTCGGGGCTTCAAGGCTGCCCCCTTGA	1001
OY	826	atagatgcacagccagcaagctcacagctcttacaacaccccaatgaacagggctctgagc	885
Db	1002	ATAGAGCCCATTCGCAACACTCAGGGCCATTCGACACACCCCAATATATATGATATGCTGGCC	1061
OY	886	tttgacccaagctactcgtgaagagaactcttcgcacatgattccagcagcaacggtgatatg	945
Db	1062	TTTCGGGACCAAGTACTCTAAGAGAGATCTTCTCCCTGGGTTCCACAGCAATGCGGATGG	1121
OY	946	gaaaaaactctggagatacacatgaagaagaagactga 984	
	1122	GAATAAATACTTGGGATCTCAATGAAGAAGTAGACTGA 1160	

RESULT	6
LOCUS	AK018579
DEFINITION	1578 bp mRNA linear HTC 19-JAN-2002
ACCESSION	AK018579
VERSION	AK018579.1
KEYWORDS	GI:12858357
SOURCE	HTC; CAP trapper.
ORGANISM	Mus musculus (strain:C57BL/6J) adult male colon CDNA to mRNA, clone.lib:RIKEN full-length enriched mouse CDNA library clone:9030625M01.
REFERENCE	Mus musculus
TITLE	Eudaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL	1 (sites)
MEDLINE	Carninci, P. and Hayashizaki, Y.
PUBMED	High-efficiency full-length cDNA cloning
REFERENCE	Meth. Enzymol. 303, 19-44 (1999)
AUTHORS	99279253
	10349636
	2 (sites)
	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

TITLE	Ichihara, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
PUBLISHED	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	2049374
AUTHORS	3 (sites)
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carninci, P.,
JOURNAL	Kono, H., Akiyama, J., Nishi, K., Katsuna, T., Teshiro, H., Itoh, M.,
PUBLISHED	Sun, N., Ishii, Y., Nakamura, S., Hazama, N., Nishine, T., Hatada, A.,
REFERENCE	Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
AUTHORS	Fujiwaka, S., Inoue, K., Togawa, Y., Iizawa, M., Ohara, E., Watanabe, K.,
TITLE	Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
JOURNAL	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
PUBLISHED	RIKEN integrated sequence analysis (RISA) system-384-format
REFERENCE	sequencing pipeline with 384 multipipillary sequencer
AUTHORS	Genome Res. 10 (11), 1757-1771 (2000)
TITLE	20350913
JOURNAL	11076861
PUBLISHED	4 (sites)
REFERENCE	The RIKEN genome exploration research group phase II team and the
AUTHORS	PANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
PUBLISHED	5 (bases 1 to 1578)
REFERENCE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
AUTHORS	Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Butt, C.,
TITLE	Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanaoka, T.,
JOURNAL	Hane, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Horii, F.,
PUBLISHED	Haru, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,
REFERENCE	Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,
AUTHORS	Kurikawa, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
TITLE	Nunakata, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
JOURNAL	Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
PUBLISHED	Schimi, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T.,
REFERENCE	Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
AUTHORS	Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yananaka, I.,
TITLE	Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
JOURNAL	Hayashizaki, Y.
PUBLISHED	Direct Submissions
REFERENCE	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The institute of
AUTHORS	Physical and Chemical Research (RIKEN), Laboratory for Genome
TITLE	Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
JOURNAL	RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
PUBLISHED	Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp,
REFERENCE	URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
AUTHORS	Fax: 81-45-503-9216)
TITLE	Please visit our web site (http://genome.gsc.riken.go.jp/) for
JOURNAL	further details.
PUBLISHED	cDNA library was prepared and sequenced in Mouse Genome
REFERENCE	Encyclopedia Project of Genome Exploration Research Group in Riken
AUTHORS	Genomic Sciences Center and Genome Science Laboratory in RIKEN.
TITLE	Division of Experimental Animal Research in Riken contributed to
JOURNAL	prepare mouse tissues. First strand cDNA was primed with a primer
PUBLISHED	[5'] GAGGAGAGATTCGATTAATTAATTAATTCACCCCCC 3']. cDNA was
REFERENCE	prepared by using trehalose thermo-activated reverse transcriptase
AUTHORS	and subsequently enriched for full-length by cap-trapper. cDNA went
TITLE	through one round of normalization to Rot -10.0 and subtraction to
JOURNAL	Rot = 185.2. Second strand cDNA was prepared with the primer
PUBLISHED	adapter of sequence 15'
REFERENCE	GAGGAGAGATTCGATTAATTAATTAATTCACCCCCC 3']. cDNA was cleaved
AUTHORS	with BamHI and XhoI. Vector: a modified pluescript KS(+) after
TITLE	bulk excision from lambda FLX I. Cloning sites, 5' end: SalI; 3'
JOURNAL	end: BamHI. Host: DH10B
PUBLISHED	Location/Qualifiers
REFERENCE	1..1578
AUTHORS	/organism="Mus musculus"
TITLE	/strain="C57BL/6J"
JOURNAL	/db_xref="MGI:MGI:1894816"
PUBLISHED	/db_xref="taxon:10090"
REFERENCE	/clone="9030625M01"
AUTHORS	/sex="male"
TITLE	/tissue_type="colon"

```

/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
216. .1202
CDS

```

```

/codon_start=1
/protein_id="BAB31290.1"
/db_xref="GI:12858358"
/translation="MCSFVYDIEDIPLEDDDPNSIEFKILAFYARHHVFNKTPAVES
PKLSRSLISOKALGTWSTPQVQVSLPCRSPPSEKKISLGKKSSSRTEPRAEKKE
EGLSSPKKEIPAAQPCQGPPEVQROSGVLSNMPKPSISVRLKSEVSDVACIAGN
VAALYISMPEDVYTHSGGSKLKRNSILILFRREGDPDSKNKGCEDIISKIYELL
FSGQLGREIKKDKALMSFDDGLSYSTFKITITLDFLDVITRGSEKAVAGFTRALL
IDAIATKLAIDNHKMRNLGFGVYLRLEFSPWQNGCWEMKILGISHEVD"
BASE COUNT      427 a      393 c      420 g      338 t
ORIGIN

```

Query Match	53.68;	Score 527;	DB 11;	Length 1578;
Best Local Similarity	74.38;	Pred. No. 2.4e-137;		
Matches 742;	Conservative	0;	Mismatches 230;	Indels 27;
				Gaps 5;

QY 1 atcgttcgacccagctgggtgtgtgacctggaagaatctccctcagatgatgatgacctaac 60
Db 216 atcgttcgacccagctgggtgtgtgacctggaagaatctccctcagatgatgatgacctaac 275

61 accatagaattcaaaatctctgcctactacacacagacatcatgtcttcaagagacccct 120

DB	4/6	ACGATAGAGTTCAAAATCTTCGGCTTTCACGCCGACACCATCTCTTCAAGAACACCCCG	335
QY	121	gcctctcttcacccaagctgcgagaacaagattgtcccaagagcgctcggggaat	180

Db 336 GGTGTCCTTTCTCGGCCAAGCTCTCCAGAACAGAGAGTCTCTCCAGAAAGCCCTGGGGACT 395

Oy 181 TATTCAACAATAGTACATGACAGAGTGTGTCATGTGCTTGCAGAAATTCCTCAATCCAGT 240

Db 396 TGGT---CAACTGATTTCTGGACACAGSTATCATGTGCTTGCAGAGGTTCCTCCAGC 452

Db 453 GAAAGACATCAGCTTGGGCAAGGAAGTCTTCTTGAGAACACTCTTCAGGGTGGCC 512

[illegible]

Oy 350 ---aaagcagctltgaataaccagaatctgcacagccagcagtcggtccagtgtcttct 405
573 cccttcccggtagagccgcatgatggctttccacaaacacacacaggccccagttctgacg 632

406 aacgtggaagcagttgtcttggagcatgaagctgtggagccccaagtcattccattggccaac 465

DB 633 AGCTGTGGAGCAGCGCCCTGCGAGAGTGAAGCTTGCGATTGCCAAGCTGGCTTTGATTTGCCAAC 632

QY 466 CGAGTAGCTGAATTTGTTATCTCTGGCCACCCACCAAGCGAGCCCGAGGAGGCTTC 525

Qy	526	aagltccaaagagatllttlgtlaactlgagggctcttccttcacagcttccaaagagccatgctt	585
Db	693	acagatggctgaaattgtttttacttcttgccaccacacagatgtcatctccacagccagggagga	752

Db 753 AGCAAGCTCAAGAGAGGGCTCTCTGGAGATTCTTACTTCAGGTTTGAGGAC-----CTT 807

Ky 366 gacgcctcaagcttcccaagaaagctgaagaagaacadaaaccaacccaadaatctgtttgagctg 645
 Db 808 GTGACTCTAAG---AATAAGATGGAGAGAGACCAATAATTAAGCAAGATTCTTGAGAGCTG 863

QY 646 ctcgaatatcttcgagagatcagttctggaagaacgtctgaagaagaataagcttgcattgcatggc 705

Db 864 cttcaaatcttctcgggggattcagtttgggagagacataaagaaagcaaacgctttgatgtgac 923,

QY 706 catccagatggtctctactctgtttccagaccatccagaccagtcctaatg 765
||||||| ||||||||| ||||||||| ||||||||| ||||| |

Db	924	AGCTTCAGAGACGGGGCTGTGCTTACTCAAGTTCAAGACCATCAGACGACGCTGTTGCTGAGG	983
QY	766	ggtgtgtgagcccccaggggagaatctcagaggttcaaaagctcaagggctttaagcttgcctctgta	825
Db	984	GACGTGTGACCCCAAGAGAGAAATCAAGAGGTCAAAAGCTCGGGGCTTTCAAGGCTGCCCTTGCA	1043
QY	826	atagaagctacagcgcaagctcacagctatttgaacaacaccggaatgaacaggttccttgagc	885
Db	1044	ATGAGAGCCCATTCGCCAAGCTCACGGCCATCGACCAACCCATGATGATGAATGCTGGGC	1103
QY	886	tttggcaccagaatcctgtaaaagagaactcttcgcataatgatccagcagcaggtgtgatgg	945
Db	1104	TTTGGGACCCAGTATCCCTAAGAGAGATCTTCTCCCTCGGTTCCAGCAGATGGCGGATGG	1163
QY	946	gaaaaaatacttggtgatatacatcatgaagaagttagactga	984
Db	1164	GAAAAATACTTGGGATCTTCACATGATAAAGATAGACTGA	1202

RESULT	7				
AK016670					
LOCUS	AK016670	2222 bp	mRNA	linear	HTC 19-JAN-2002
DEFINITION	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933405K19:BCL2-like apoptosis inhibitor containing protein, full insert sequence.				
ACCESSION	AK016670				
VERSION	AK016670.1	GI:12855533			
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,				

ORGANISM	REFERENCE
<i>Mus musculus</i>	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.	
1 (sites)	
Carninci, P. and Hayashizaki, Y.	
Title	High-efficiency full-length cDNA cloning
Journal	Meth. Enzymol. 303, 19-44 (1999)
Medline	99279253
Pubmed	10349636
Reference	2 (sites)

AUTHORS Carinciuc, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kato, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (sites)
AUTHORS	Shibatata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Saseki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunaga,S., Kawai,J., Okazaki,Y., Mutamatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (sites)
AUTHORS	The Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 2222)
AUTHORS	Adachi, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukushiki, Y., Furuno, M., Hanagaki, T., Hara, A., Haratsu, N., Hill, D., Hiromoto, K., Hirooka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S.,

KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 715)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLM1560 row: 1 column: 20
High quality sequence stop: 715.
Location/Qualifiers
1. 715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5222875"
/clone_lib="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector: pcMV-SPORE6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of spleen and pancreas from 28 yo male. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH-MGC Library."
BASE COUNT 212 a 177 c 171 g 155 t
ORIGIN

Query Match 51.5%; Score 506.6; DB 10; Length 715;
Best Local Similarity 99.2%; Pred. No. 9.9e-132;
Matches 509; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

61 atgtgtagcaccagtggtgtgtgacatggaagaatccccctagatgatgacctaac 60
203 ATGTGTAGCACCAGTGGTGTGACCTGGAAGAAATCCCTGATGATGATGACCTAAAC 262
61 accatagaattcaaatcctcgtcctactacacacacacatcatgtcttcaagagaccct 120
263 ACCATAGAATTCAAAATCCTGCTACTACACACACATCATGTCTTCAAGACACCCCT 322
121 gctcctctccacccaagctgtgagaaagaattgtccacagagggccctgggaat 180
323 GCTCTCTTCCACCAAGCTGTCTGAGAAAGAAATTTGTCACAGGGGCTGGGGAAT 382
181 tgttcagcaaatgtatgacagaggtgtcatgacctgtgagaatcccccaatcagt 240
383 TGTTCACCAATGTATGATGACAGAGGTGTCTATGCTGTGAGAAATTCCTCAATCCAGT 442
241 gagaagggcaataacacctgtgcaagaagaagcttcttctggaagcatcttggagtagtg 300
443 GAGAAGGGCCATTAACCTTGTGCAAGAAAGCTTCTTGGAAGCATTTCTTGAGAGTAGTG 502
301 gagaaggaagatctgcagagagagccttcgaagctcgtcgtcgaaggttaagaagagcttg 360
503 GAGAAGGAAGATTTGCGAGAGCAGCGCTCCCAAGGTCTGCTCAGGGGTCAAAAGGAGCTTG 562
361 gaatacaagaatctgcacagcagcagtggttccagtgcttcttcttaacgtggaagctgc 420
563 GAATACCAAGATTTGCGACAGCAGCAGTGTGCTTCTTCAACGTGAGCAGCAGTGC 622
421 ttggagcatgaagctgtggaaccccaagatcattccattgccaacccagtaagctgaatt 480

|||||
Db 623 TTGGAGCATGAGAGCTGTGACCCCAAGTCATTTCCATTGCCAACCGAGTAGTGAATC 682
Qy 481 gttattcctggccaccaccacgaagcagccag 513
|||||
Db 683 GTTTACTCTGTGACACCAACACAGCAGCAGCAG 715

RESULT 9
BG717835 672 bp mRNA linear EST 08-MAY-2001
LOCUS 602693855P1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4826292 5',
DEFINITION mRNA sequence.
ACCESSION BG717835
VERSION BG717835.1 GI:13997022
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 672)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: LLM10740 row: m column: 13
High quality sequence stop: 648.
Location/Qualifiers
1. 672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4826292"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3', size-selected for average insert size 2.2 kb and normalized to RQF 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 197 a 162 c 171 g 142 t
ORIGIN

Query Match 50.5%; Score 497.2; DB 10; Length 672;
Best Local Similarity 99.2%; Pred. No. 4.3e-129;
Matches 510; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

1 atgtgtagcaccagtggtgtgtgacatggaagaatccccctagatgatgacctaac 60
Db 159 ATGTGTAGCACCAGTGGTGTGACCTGGAAGAAATCCCTGATGATGATGACCTAAAC 218
61 accatagaattcaaatcctcgtcctactacacacacacatcatgtcttcaagaagaccct 120
219 ACCATAGAATTCAAAATCCTGCTACTACACACACATCATGTCTTCAAGAGCACCCCT 278
121 gctcctctccacccaagctgtgagaaagaattgtccacagagggcctgggaat 180
Db 279 GCTCTCTTCCACCAAGGCTGTGAGAAAGAAATTTGTCACAGGGGCTGGGGAAT 338
181 tgttcagcaaatgtatgacagaggtgtcatgacctgtgagaatcccccaatcagt 240

|||||
 Db 339 TGTTCGCAAAATAGTCATGACAGAGTGTCAAGCCTTCAGAAATCCCAATCCACT 398
 OY 241 ggaagagccataaaccttggcaagaaagctctcttggaaagcattcttggatgaag 300
 Db 399 GAGAGGCCATTAACCTTGGCAGAAAAAGCTCTTGGAAAGCATCTTTGGAGTAGTG 458
 OY 301 gagaaggaagatcgcagagcaagcctgcacaaaggtctcttgcaggtgcaaaagagcgttg 360
 Db 459 GAGAGGAGAGATTCGAGAGCAGCCTGCGCAAGGTCTCTGCTCAGGGTCAAGAGAGAGCTTG 518
 OY 361 gaataccaa-gatcgcagacagcagcagtggttccaggtgtcttcttaacgttgagacagtg 419
 Db 519 GAATACCAACGATTCGACAGCAGCAGTGTCTCAGGTCTTTCTTAACGTGAGCAGAGTG 578
 OY 420 ctggagagcagtagcctgtggaccacaaagatcattccttgcacacagagtagctgaagt 479
 Db 579 CTGGAGCATGAAAGCTGTGGACCCCAAGTCATTTCCTCCAAACCGAGTAGCTGAAT 638
 Db 480 tgttatctctggccacacacacagcagaccag 513
 Db 639 TGTATTCCTCTGTGCAACACACACAGGAGACAG 672

RESULT 10

BI908814 703 bp mRNA linear EST 16-OCT-2001
 LOCUS BI908814 60306627F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5215218 5',
 DEFINITION MRNA sequence.
 ACCESSION BI908814
 VERSION BI908814.1 GI:16171889
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 703)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LAM11540 row: J column: 19
 High quality sequence stop: 701.
 Location/Qualifiers

FEATURES

source

1..703
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5215218"
 /clone_lib="NIH_MGC_118"
 /tissue_type="Leukocyte"
 /lab_host="DH10B"
 /note="vector: PCMV-SPORT6, site.1: NotI, site.2: EcoRV (destroyed); RNA source leukocytes from anonymous pool of non-activated adult donors. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH_MGC Library."
 BASE COUNT 208 a 173 c 170 g 152 t
 ORIGIN

Query Match 42.5%; Score 418.4; DB 10; Length 703;

Best Local Similarity 98.3%; Pred. No. 7,le-107;
 Matches 507; Conservative 0; Mismatches 1; Indels 8; Gaps 8;

OY 1 atgtgtgaccacagaggtgtgactctggaagaatccccctgatatgatgtgacctaacc 60
 Db 190 ATGTGTACACCAAGTGGGTGTGACTGTGAGAAATCCCTTAATGATGTGACCTTAAC 249
 OY 61 accataagatccaanaatctcgcctactacacacagac-atcatgtcttcaagacacccc 119
 Db 250 ACCATFAGAAATTCAAATCTCTGCTACTAGACAGACTATCATGTCTTCAAGAGACACCC 309
 OY 120 tgcctctcttcacccaagcctgtctgagaacaaagattgttccagaagggtctgggaa 179
 Db 310 TGCTCTCTTCTCACCAAGACTGCTGAGAACAAAGATTGTGCCAAGAGGGGCTGGGAA 369
 OY 180 ttgttcgcaaatgatagtatgacagaggtgtgtatgaccttggaaatctccaatccag 239
 Db 370 TTTGTTCAACCAATGAGTATGAGACAGAGGTGTATGGCTTTGCAAGAAATCCCAATCCAG 429
 OY 240 tgaagaagccataaaccttggcaagaa-aaagtctcttggaaagatctct-tggagta 297
 Db 430 TGAGAAGGCCATTAACCTTGGCAAGACAAAGTCTTGTGGAAGCATTTGTGAGTA 489
 OY 298 gtggaagaagaagatctg-cagagcaagcctgcgaagctctctgtcaggtgcaagagac 356
 Db 490 GTGGAGAAAGGAAGATTCCGCCAGACAGCAGCTGCCAAGGTCTCTGCTCAGAGTCAAGAGAC 549
 OY 357 gtgtgaatcaccaagatctgcacagc-cagcaatgtgtccaggtgtcttcttaa-cgtggag 414
 Db 550 GTTGAATACCAAGATTGTGACAGCGCACAGTGTCTCAGGTCTTTTAACTGGAG 609
 OY 415 cagtccttgagcatgaagcctgtgaccccaagatcattcattgcacacagtagct 474
 Db 610 CAGTGTCTTGAGCATGAGCTGTGACCCCAAGTATTTCCATTTCACCAAGCAGTAGCT 669
 OY 475 gaattgttatctcctggccacacacacagcgacc 510
 Db 670 GAAA-TGTTATTCCTGTGCA-CACACACAGGACG 703

RESULT 11

AW000827/c 426 bp mRNA linear EST 09-MAR-2000
 LOCUS AW000827 w445d10.x1 Soares.Dieckgraefe.colon_NHCD Homo sapiens cDNA clone
 DEFINITION IMAGE:2522995 3', mRNA sequence.
 ACCESSION AW000827
 VERSION AW000827.1 GI:5847665
 KEYWORDS EST.

SOURCE

human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 426)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 634 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 419.
 Location/Qualifiers

FEATURES

source

1..426
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2522995"
 /clone_lib="Soares.Dieckgraefe.colon_NHCD"
 /tissue_type="colonic mucosa from 3 patients with Crohn's
 disease"
 /lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Tissue samples provided by Dr. Brian Dieckgraefe (Washington University, dieckgrbm.wustl.edu); colonic mucosa represents a range of disease involvement from moderate to severe Crohn's disease; samples include both perforating (fistulas) and non-perforating samples. Library constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 91 a 108 c 88 g 139 t

Query Match 38.2%; Score 376; DB 9; Length 426;
Best Local Similarity 99.3%; Pred. No. 5.2e-95;
Matches 399; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

584 ctgtagctcaagtcctcaagaagatgaagaacaatactag-ccaaattgttgag 642
|||||
426 CTGTAGCTCAAGTCTTCAAGAAAGTGAAGAAACAATACTAGCCCAAAATTGTTGAG 367
643 ctgctgaatattcaggagatcagcttggaagaagctgaagaataagcttgatg 702
|||||
366 CTGCTGAATATTTCAGAGATCAGTGTGAAGAAAGC--AAGAAGATTAAGCTTTGATG 309
703 ggcacttcaggatgggtgtcctactctgtttcaagacatcacagaccagttccta 762
|||||
308 GGCACACTCCAGATGGGCTGTCTACTCTGTTTCAAGACCATCACAGCCAGTCCCTA 249
763 atgggtgtggaaccccaaggaggatcaaggtcaaggtcagggttgaagtgccctt 822
|||||
248 ATGGGTGTGACCCCAAGGGAGATCAAGGTCAAGGCTCAAGGGCTTTAAGGCTCCCTT 189
823 gtaataagctcaagcgaagctcacagctatgacaacaccacgcgaatgaagaagttcctg 882
|||||
188 GTATAGACCTCAGCGCCAACTCAGCTATTGACAAACCCGATGAACAGGTCCTG 129
883 ggccttggcaccagatcctcgaagaagaactctgcgcattgcatcagagcaggtgga 942
|||||
128 GGCCTTGGCACCAAGTACTGAAGAGAACTCTCGCCATGATCAGAGCAGCGTGA 69
943 tgggaataaatacttggtgatatacatalgaagaagtagactga 984
|||||
68 TGGGAAAAAATCTTGGGATATCATGAAAGATGAGACTGA 27

RESULT 12
BI652080 676 bp mRNA linear EST 12-SEP-2001
LOCUS BI652080
DEFINITION 60329886F1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5339464 5',
mRNA sequence.
ACCESSION BI652080
VERSION BI652080.1 GI:15566316
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; 1 (bases 1 to 676)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: LHM11862 row: c column: 17
High quality sequence stop: 676.
Location/Qualifiers

FEATURES

source

1..676

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5339464"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert 2 kb. Library constructed by Life Technologies, catalog #12017-018. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999). Note: this is a NIH CGAP Library."

BASE COUNT 194 a 156 c 189 g 137 t

Query Match 36.5%; Score 359; DB 10; Length 676;
Best Local Similarity 78.9%; Pred. No. 3.9e-90;
Matches 482; Conservative 0; Mismatches 115; Indels 14; Gaps 4;

376 caccagccagatggtccaggtgtcttctcaagtgaagcaggtccttggagcagaact 435
|||||
16 CACAACACACACTGGGCCAGGTCTGAGCAGTGTGAGCAGCGCTGGAGAGTGAAGTT 75
436 gggagcccaagatccatccatgccaacccagtagctgaattgtttatcctggcca 495
|||||
76 GTGGATTCCAAAAGTGGCTTGTATTGCCAACAGAGTGGCTGAATTTACTCCGGCCA 135
495 gggagcccaagatccatccatgccaacccagtagctgaattgtttatcctggcca 495
|||||
496 ccaccacaagcga--ccagggcaggaggtcctcaagtcacaagaagatttgtaactgagg 553
|||||
136 CCACCAAGATGTATCCACAGCCAGAGAGAGCAAGCTCAAGAGAGA--GGGTCTCGAGA 192
554 gctctccttccagctcccaagcccaagctgctgtgacttcaagttcgaagaagtgaag 613
|||||
193 TTCTGTACTTCAAGGTTGAAGGAC---CTTGTGACTCTTAAG---AATTAAGATGAG 243
614 aagaacaatactagcacaataattgttagctgtcgaatattcaggaagatcaattggaaa 673
|||||
244 AAGACCAAAATTAATCAACAGATGTTAGCTGCTGAATTTCTCGGGGATCAGTTGGGAA 303
674 gaaagctgaagaagaataaggtcttgatggccaacttccaggaatgggtgtcctactctg 733
|||||
304 GAGAGATTAAGAAAGACAGAGGCTTTGATGAGACAGCTTCCAGAGCGGGCTGTCTACTCA 363
734 ttccaagaccatcaagaccaggtctctaattggtgtggtgaccccaaggagggaataagagg 793
|||||
364 CGTTCAAGACCAATCAAGACCTGTCTGTAGGGAGCGTGGACACACAGAGAAATATGAGG 423
794 tcaagctcaaggtcttaagctgtgccccttgtaataagacgtcaagcgaagctcaagccta 853
|||||
424 TCAAGAGCTGGGGCTTCAAGGCTGCGCTTGCATATAGACGCCATGCCAAGCTCAGGGCCA 483
854 ttgacaacaccaccogattgaacaaggtctcctggcttggcaccagaatlaacctgaagaagaact 913
|||||
484 TCGACCAACCAACCAATAGATAGATGCTGGGCTTCGGGACCAATACTAGAGAGTACT 543
914 ttctgcacatggatccagcagcagcaggtgagatggggaataaatacttgggtatcacatgtag 973
|||||
544 TCTCCCTGCTGCTTCAGCAGAAATGCGGATGGGAAAAAATATCTGGGATCTCATATGAG 603
974 aagtagactga 984
|||||
604 AAGTAGACTGA 614

```

RESULT_13
LOCUS      AQJ31601                      422 bp          DNA           linear       GSS-23-SEP-1998
DEFINITION HS_3049_A1.D07_MR.CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3049 COL=13 Row=G, DNA sequence.
ACCESSION  AQJ31601
VERSION    AQJ31601.1   GI:3508767
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 422)
AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J.J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE      Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL    Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENTARY Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel.: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Sequence Tagged Connector
            Plate: 3049 row: G column: 13
            Class: BAC ends
            High quality sequence stop: 422.
FEATURES   Location/Qualifiers
             source              1..422
                                /organism="Homo sapiens"
                                /db_xref="taxon:9606"
                                /clone_id="Plate=3049 COL=13 Row=G"
                                /clone_lib="CIT Approved Human Genomic Sperm Library D"
                                /sex="male"
                                /note="Organ: sperm; Vector: pBelOAc11; BAC clones In E-Coli DH10B"
BASE COUNT      110 a        101 c        104 g        105 t         2 others
ORIGIN
Query Match               36.3%; Score 356.8; DB 12; Length 422;
Best Local Similarity     95.1%; Pred. No. 1.3e-89;
Matches 367; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

```

JOURNAL LINE COMMENT

68	aattcaaaatcttcgctactacaccgaacatcatgttctaagagcacccttgtctct	127
Dd	ATTTCAAAATCTTCGCGCCATCACACAACATGAAGTGTTAAGAAGCACCCCTGCCTCT	95
QY	tctaccaagaacctcttgtagacaagaagtgtcccgaggggccttggggaattttcag	187
Dd	TCTGACCCTACTCTGCTGGAAACAAGAAATTGTGCCACAGGGGCTCGCGAATTITTGAG	155
QY	188 caaatgagtcattgacagaggtgtcatctgccttgcgaagaatcccacatccagtgaaga	247
Dd	CATPMTAGTCAATGACAGAGGTGTCATGCGCTTGCAAGAAATCCCACATCCAGTAGAAGG	215
QY	248 ccataaaccttggccaagaaaaagtctcttcttggaaaagcattcttggagtaigtgaaga	307
Dd	CCATPAAACCTTTGGCAAAGAAAAGTCCTTTGTGGAAAAAGCATTCCTTTGAGTAGTGAAGG	275
QY	308 aagaatgccagagcacggctccaaagrtcttgctccaggrccaagggacgttggatatcc	367
Dd	AAGATTGCGACAAGCACGGCTGCCAAGGCTCTGCTGCACANGTCCAAGAGCAAGCTTGGAATACC	335
QY	368 aagaatgccagagcacggagrgtltccaaigtgttctttaacgttggagagtgcttggagc	427
Dd	AAGATTGCGACAAGCACAGCAAGTAGTGCCATGTCTTCCTTCAAGCTGAGCAGTAGCTTGAGC	395
QY	428 atgaagctgttggaccccaagtcatt	453

DB	396	ATGACGTANGCATCTCGGATTCITT	421
RESULT	14		
LOCUS	BB013694		
DEFINITION	BB013694 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4930470009 3', mRNA sequence.		
ACCESSION	BB013694		
VERSION	BB013694.2		
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.		
AUTHORS	1 (bases 1 to 984) Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takehashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.		
TITLE	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)		
JOURNAL	Unpublished (2001)		
COMMENT	On Jun 1, 2000 this sequence version replaced gi:8136849. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Itoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1571-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saio,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.		
FEATURES			
SOURCE			
	1..984		
	/organism="Mus musculus"		
	/strain="C57BL/6J"		
	/db_xref="taxon:10090"		
	/clone="4930470009"		
	/clone_lib="RIKEN full-length enriched, adult male testis (DH10B)"		
	Location/Qualifiers		

Db 292 GTGAGAGGCCATMAACCTTGCAAGAAAAAGTCTTCTTGAAAGCATTCCTTGAGTAG 351
Oy 299 tggagaaagaaatctgcagagcacgcctgcgaaggtctctgcacaggtcaaaag 353
|||||
Db 352 TGGAGAGGAGAGATTGCGAGAGCACGCTTGCCAAAGGTCTTGCTCAGGGTCAAG 406
|||||

Search completed: June 20, 2002, 04:19:16
Job time: 6406 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2002, 16:33:47 ; Search time 15.65 Seconds
(without alignments)
809.028 Million cell updates/sec

Title: US-09-771-961-2
Perfect score: 1709
Sequence: 1 MCSTSGCDLEIPDDDLN.....WIOHGGWEKILGISHEVD 327

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	128.5	7.5	239	1	BCLX.CHICK
2	119	7.0	233	1	BCLX.MOUSE
3	118	6.9	233	1	BCLX.RAT
4	117	6.8	233	1	BCLX.HUMAN
5	115	6.7	530	1	TDT.MOUSE
6	114	6.7	233	1	BCLX.PIG
7	105.5	6.2	901	1	SECA.HARTN
8	99	5.8	757	1	YVDK.BACSU
9	97	5.7	537	1	SYK.AQDAE
10	96	5.6	2319	1	AKA6.HUMAN
11	94.5	5.5	192	1	BAXA.MOUSE
12	93	5.4	204	1	AR11.XENLA
13	93	5.4	859	1	PMS2.MOUSE
14	92.5	5.4	723	1	GGA3.HUMAN
15	92.5	5.4	945	1	RPN2.YEAST
16	92	5.4	880	1	DPO1.BACSU
17	91.5	5.4	520	1	TDT.BOVIN
18	91	5.3	875	1	Z1P1.YEAST
19	90.5	5.3	192	1	BAXA.RAT
20	90	5.3	433	1	KAR1.YEAST
21	89.5	5.2	218	1	BAXB.HUMAN
22	89.5	5.2	696	1	UVRG.DROME
23	89.5	5.2	2319	1	FAB.MOUSE
24	89	5.2	926	1	POOL.HAELN
25	89	5.2	1732	1	RBB2.HUMAN
26	88.5	5.2	445	1	FLDL.VIRPA
27	88.5	5.2	2524	1	NOTC.XENLA
28	88.5	5.2	3027	1	POLG.PYFV1
29	88	5.1	196	1	TECI.YEAST
30	87.5	5.1	482	1	BAXA.HUMAN
31	87	5.1	1	1	TDT.MONDO
32	87	5.1	854	1	AKA1.RAT
33	87	5.1	947	1	TOP1.MYCLE

34	86.5	5.1	1165	1	SYV.AQDAE	067411 aquifex aeo
35	86.5	5.1	2875	1	RPL7.TSMV1	P28976 tomato spot
36	86	5.0	537	1	IMDH.DROME	007152 drosophila
37	86	5.0	856	1	CLPB.HELPI	P71404 helicobacte
38	85.5	5.0	250	1	YC97.HUMAN	Q9Y345 homo sapien
39	85.5	5.0	1158	1	R114.HUMAN	P48552 homo sapien
40	85	5.0	229	1	BCL2.BOVIN	002718 bos taurus
41	85	5.0	239	1	BCL2.HUMAN	P10415 homo sapien
42	85	5.0	639	1	GGA1.HUMAN	09U355 homo sapien
43	85	5.0	702	1	DY12.ANTCR	016959 antiochidari
44	84.5	4.9	192	1	BAXA.BOVIN	002703 bos taurus
45	84.5	4.9	697	1	YE9C.SCHPO	013773 schizosacch

ALIGNMENTS

RESULT	ID	STANDARD	PRT	AA
1	BCLX.CHICK			
BCLX.CHICK	007816; Q98908;			
01-FEB-1995 (Rel. 31, Created)				
01-NOV-1997 (Rel. 35, Last sequence update)				
16-OCT-2001 (Rel. 40, Last annotation update)				
Apoptosis regulator Bcl-X.				
BCL2L1 OR BCLX OR BCL-X.				
Gallus gallus (Chicken).				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
Gallus.				
NCBI_TaxID=9031;				
SEQUENCE FROM N.A. (SHORT FORM).				
MEDLINE=93364977; PubMed=8358789;				
Boise L.H., Gonzalez-Garcia M., Postema C.E., Ding L., Lindsten T.,				
Turka L.A., Mao X., Nunez G., Thompson C.B.;				
"bcl-x", a bcl-2-related gene that functions as a dominant regulator				
of apoptotic cell death.";				
Cell 74:597-608(1993).				
(2)				
SEQUENCE FROM N.A. (LONG FORM).				
STRAIN=HUBBARD WHITE MOUNTAIN; TISSUE=Testis;				
RA				
Vilagrassa X., Mezquita C., Mezquita J.;				
"differential expression of bcl-2 and bcl-x during chicken				
spermatogenesis.";				
Mol. Reprod. Dev. 47:26-29(1997).				
-1- FUNCTION: DOMINANT REGULATOR OF APOPTOTIC CELL DEATH. THE LONG				
FORM DISPLAYS CELL DEATH REPRESSOR ACTIVITY, WHEREAS THE SHORT				
ISOFORM PROMOTES APOPTOSIS (BY SIMILARITY).				
-1- SUBCELLULAR LOCATION: MITOCHONDRIAL MEMBRANES AND PERINUCLEAR				
ENVELOPE (BY SIMILARITY).				
-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A				
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.				
-1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN ORGANS WITH LYMPHOID				
DEVELOPMENT.				
-1- DOMAIN: BH4 DOMAIN SEEMS TO BE INVOLVED IN THE ANTI-APOPTOTIC				
FUNCTION. INTACT BH1 AND BH2 DOMAINS ARE REQUIRED FOR ANTI-				
APOPTOTIC ACTIVITY (BY SIMILARITY).				
-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 1 (BH1).				
-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 2 (BH2).				
-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 3 (BH3).				
-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 4 (BH4).				
-1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.				

This SWISS-PROT entry is copyright. It is produced through a collaboration				
between the Swiss Institute of Bioinformatics and the EMBL outstation -				
the European Bioinformatics Institute. There are no restrictions on its				
use by non-profit institutions as long as its content is in no way				
modified and this statement is not removed. Usage by and for commercial				
entities requires a license agreement (See http://www.isb-sib.ch/announce/				
or send an email to license@sib-sib.ch).				

```

DR EMBL: U01010; AAA82172.1; -.
DR EMBL: U51279; AAC53460.1; -.
DR EMBL: U78031; AAB96881.1; -.
DR EMBL: U78030; AAB96881.1; JOINED.
DR HSSP: P53563; IAF3.
DR MGI: MGI:88139; Bcl2l.
DR InterPro: IPR002475; BCL2_family.
DR InterPro: IPR003093; BH4.
DR InterPro: IPR00712; BCL_2.
DR Pfam: PF00452; Bcl-2; 1.
DR Pfam: PF02180; BH4; 1.
DR SMART: SM00337; BCL; 1.
DR SMART: SM00265; BH4; 1.
DR PROSITE: PS50062; BCL2_FAMILY; 1.
DR PROSITE: PS01080; BH1; 1.
DR PROSITE: PS01258; BH2; 1.
DR PROSITE: PS01259; BH3; 1.
DR PROSITE: PS01260; BH4_1; 1.
DR PROSITE: PS50063; BH4_2; 1.
DR Apoptosis; Mitochondrion; Alternative splicing; Transmembrane.
FT DOMAIN 4 24 BH4.
FT DOMAIN 86 100 BH3.
FT DOMAIN 129 148 BH1.
FT DOMAIN 180 195 BH2.
FT TRANSMEM 210 226 POTENTIAL.
FT VARSPLIC 126 188 MISSING (IN ISOFORM BCL-X(S)).
FT VARSPLIC 189 233 DFEVDLVGNNAAESRKGQEFNFRLFMFLGMVAGVLLGSL
FSRK -> VRTPLVCPPLACVSLCEHP (IN ISOFORM
BCL-X(BETA)).
FT VARSPLIC 194 233 LYGNNAAESRKGQEFNFRLFMFLGMVAGVLLGSLFSRK
-> GHDCGKGSGAGLTLOSEVTRH (IN ISOFORM BCL-
X(DELTA-TM)).
SO SEQUENCE 233 AA; 26132 MW; 24D2AC79887E072E CRC64;

Query Match 7 0%; Score 119; DB 1; Length 233;
Best Local Similarity 21.1%; Pred. No. 0.033;
Matches 54; Conservative 37; Mismatches 91; Indels 74; Gaps 11;

QY 77 SOSSEKAI-----NLGKKSSKAFGVYKEKDSQSTPAKVSAGQRTLEYQDSSHQOM 130
DB 2 SQSNRLVYDFLSYKLSQKQYVSWQSFQDVEENR---TEAPRETEAKERTPAININPWS 57
QY 131 SRCLSNVEQCLEHBAVDPKYIISANRVAEIVSWPPQATGAGFKSEIEFTGSLFQSL 190
DB 58 -----HLADSPAY-----NGAT----- 69
DB 191 OGHVPAVSSSKKDEEEDL-LAKIVELLYSGDLE---RLKKKKALMGHPDGLSYSVF 246
DB 70 -GH-----SSLDAREVTPMAAVKQALREADEDEFLRYRRAFSDLTSLHTPGTAQSF 123
QY 247 KTTIDQVLM-GVDPGRGESEVKAQFKALVLDVYAKLTAIDNHPNRRVYLGFGTKYLLKENF 305
DB 124 EQVYNLFLFRGVV--WRIYAFVFFSGALCVESYDKEMQV---LVSRLASMMATYINDHL 178
QY 306 SEWIOGHGMEKILGT 321
DB 179 EPWIDENGWDFEVDL 194

RESULT 3
BCLX_RAT STANDARD; PRT; 233 AA.
AC P53563; Q62678; P70614; P70613; Q62836; Q64087; Q64128;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Apoptosis regulator Bcl-x.
GN BCL2L1 OR BCL2L OR BCLX.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxId=10116;

```

```

RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS X(L) AND X(S)).
RC TISSUE=Brain;
RA Michaelidis T.M.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Messelinger S.L., David G.L., Choi S., Velluona M., Hardwick J.M.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS X(L) AND X(BETA)).
RC TISSUE=Thymus;
RX MEDLINE=96278736; PubMed=8662675;
RA Shiraawa N., Inohara N., Okada S., Yuzaki M., Shoji S.-I., Ohta S.;
RT "An additional form of rat Bcl-x, Bcl-xbeta, generated by an
unspliced RNA, promotes apoptosis in promyeloid cells.";
RL J. Biol. Chem. 271:13236-13265(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS X(L) AND X(S)).
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=ovary;
RX MEDLINE=95129487; PubMed=7828536;
RA Tilly J.L., Tilly K.I., Kenton M.L., Johnson A.L.;
RT "Expression of members of the bcl-2 gene family in the immature rat
ovary: equine chorionic gonadotropin-mediated inhibition of granulosa
cell apoptosis is associated with decreased bax and constitutive
bcl-2 and bcl-xlong messenger ribonucleic acid levels.";
RL Endocrinology 136:232-241(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=9801630; PubMed=9346936;
RA Artom M., Kunishima N., Inohara N., Ishibashi Y., Ohta S.,
RA Morikawa K.;
RT "Crystal structure of rat Bcl-XL. Implications for the function of
the Bcl-2 protein family.";
RL J. Biol. Chem. 272:27885-27892(1997).
CC -1- FUNCTION: Potent inhibitor of cell death. Inhibits activation of
caspases (By similarity). Appears to regulate cell death by
blocking the voltage-dependent anion channel (VDAC) by binding
to it and preventing the release of the caspase activator,
cytochrome c, from the mitochondrial membrane. The Bcl-x(s) and
Bcl-x(beta) isoforms promote apoptosis.
CC -1- SUBUNIT: Bcl-x(l) forms heterodimers with BAX, BAK and Bcl-2 (By
similarity). Heterodimerization with BAX does not seem to be
required for anti-apoptotic activity (By similarity).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MEMBRANES AND PERINUCLEAR
ENVELOPE (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; BCL-X(L) (SHOWN HERE), BCL-X(S)
AND BCL-X(BETA); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES. BCL-X(BETA) IS
SPECIFICALLY EXPRESSED IN CEREBELLUM, HEART, AND THYMUS. IN THE
OVARY, THE PREDOMINANT FORM IS BCL-X(L), WITH A SMALL BUT
DETECTABLE LEVEL OF BCL-X(S).
CC -1- DOMAIN: The BH4 domain is required for anti-apoptotic activity.
The BH1 and BH2 domains are required for both heterodimerization
with other Bcl2 family members and for repression of cell death.
CC -1- PTM: Proteolytically cleaved by caspases during apoptotic
cleaved protein, lacking the BH4 domain, has pro-apoptotic
activity (By similarity).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOGY DOMAIN 1 (BH1).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOGY DOMAIN 2 (BH2).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOGY DOMAIN 3 (BH3).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOGY DOMAIN 4 (BH4).
CC -1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```


CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; BCL-X(L) (SHOWN HERE), BCL-X(S)
CC AND BCL-X(BETA); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: BCL-X(S) IS EXPRESSED AT HIGH LEVELS IN CELLS
CC THAT UNDERGO A HIGH RATE OF TURNOVER, SUCH AS DEVELOPING
CC LYMPHOCTES. IN CONTRAST, BCL-X(L) IS FOUND IN TISSUES CONTAINING
CC LONG-LIVED POSTMITOTIC CELLS, SUCH AS ADULT BRAIN.
CC -1- DOMAIN: The BH4 domain is required for anti-apoptotic activity.
CC The BH1 and BH2 domains are required for both heterodimerization
CC with other Bcl2 family members and for repression of cell death.
CC -1- PTM: Proteolytically cleaved by caspases during apoptosis. The
CC cleaved protein, lacking the BH4 domain, has pro-apoptotic
CC activity.
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 1 (BH1).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 2 (BH2).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 3 (BH3).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 4 (BH4).
CC -1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z23116; CAAB0662.1; -
CC EMBL; Z23115; CAAB0661.1; -
CC EMBL; U72398; AAB17354.1; -
CC PDB; 1BXL; 29-OCT-97.
CC PDB; 1LXL; 21-APR-97.
CC PDB; 1MAZ; 21-APR-97.
CC MIM; 600039; -
CC InterPro: IPR002475; BCL2_family.
CC InterPro: IPR003093; BCL2.
CC InterPro: IPR000712; BCL_2.
CC Pfam; PF00452; BCL2; 1.
CC Pfam; PF02180; BH4; 1.
CC SMART; SM00337; BCL; 1.
CC SMART; SM00265; BH4; 1.
CC PROSITE; PS50062; BCL2_FAMILY; 1.
CC PROSITE; PS01080; BH1; 1.
CC PROSITE; PS01258; BH2; 1.
CC PROSITE; PS01259; BH3; 1.
CC PROSITE; PS01260; BH4_1; 1.
CC PROSITE; PS50063; BH4_2; 1.
CC Apoptosis; Mitochondrion; Alternative splicing; Transmembrane;
CC 3D-structure.
CC DOMAIN 4 24 BH4.
CC FT DOMAIN 86 100 BH3.
CC FT DOMAIN 129 148 BH1.
CC FT DOMAIN 180 195 BH2.
CC FT TRANSMEM 210 226 POTENTIAL.
CC FT SITE 61 61 CLEAVAGE BY CASPASE-1.
CC FT VARSPLIC 126 188 MISSING (IN ISOFORM BCL-X(S)).
CC FT VARSPLIC 189 233 DFEVELIGNNAASRQGFNRFLTGMTAGVVLGSL
CC FT VARSPLIC 189 233 FSRK -> VRTKPLVCFSLSGRSPALLYLFLCWIV
CC FT VARSPLIC 189 233 VGDVDS (IN ISOFORM BCL-X(BETA)).
CC FT MUTAGEN 61 61 D->A: NO CLEAVAGE BY CASPASE-1 NOR BY
CC FT MUTAGEN 131 133 CASPASE-3.
CC FT MUTAGEN 135 137 FRD->VRA: NO HETERODIMERIZATION WITH BAX.
CC FT MUTAGEN 138 140 VNM->ALL: LOSS OF ANTI-APOPTOTIC
CC FT MUTAGEN 138 140 ACTIVITY.
CC FT MUTAGEN 138 140 GRI->ELN: LOSS OF ANTI-APOPTOTIC
CC FT MUTAGEN 138 140 ACTIVITY.
CC FT MUTAGEN 138 140 G->A: NO HETERODIMERIZATION WITH BAX.
CC FT MUTAGEN 138 140 G->E: NO HETERODIMERIZATION WITH BAX.
CC FT MUTAGEN 136 156 D->A: NO EFFECT ON CASPASE-1 CLEAVAGE.
CC FT MUTAGEN 176 176 D->A: NO EFFECT ON CASPASE-1 CLEAVAGE.
CC FT MUTAGEN 188 189 WD->GA: REDUCES ANTI-APOPTOTIC ACTIVITY
CC FT MUTAGEN 189 189 BY ABOUT HALF.
CC FT MUTAGEN 189 189 D->A: NO EFFECT ON CASPASE-1 CLEAVAGE.
CC FT CONFLICT 70 70 G -> A (IN REF. 1; CAAB0661).

SQ SEQUENCE 233 AA; 26049 MW; E09D3CDD851AE9BE CRC64;
Query Match 6.8%; Score 117; DB 1; Length 233;
Best Local Similarity 20.8%; Pred. No. 0.047;
Matches 54; Conservative 35; Mismatches 80; Indels 90; Gaps 12;
QY 77 SQSSSKAI-----NLGKRSKRAFFGVVER-----EDSSTPAKVASQORTLEY 122
DB 2 SQSNELVDFLSTYLSQGYGYSMSQSDVEENTEAPEGESEMETSAING----- 53
QY 123 QDSHSQMSRCLSNVEQCLEHEAVDPKVISIANRVAEIVYSWPPQATQAGFSKELFV 182
DB 54 ---NPSW-----HLADSPAV----- 65
QY 183 TEGSLFQLOGHVPVASSSKKDEEQL-LAKIVELLYSGDQLF---KKLKRAKLMGHQ 238
DB 66 -NGAT-----GH-----SSSLDAREVIPMAAVKQALREAGDEFELRYRARSFSLTSOLHT 115
QY 239 DGLSYVEKTIIDQYILM-GVDPREGESEVKAQGFKAALVIDVTAKLTIDNHPMNRVLGFG 297
DB 116 PGTAFQSFQVYNELFRQGVN--WGRIVAFSFGALCVESVDKENVY---LVSRIAMM 170
QY 298 TKYKENFSPIQOHGWE 316
DB 171 ATYLDHLEPWIOENGWMD 189
RESULT 5
ID TDT_MOUSE STANDARD; PRT: 530 AA.
AC P09838; Q99PD0; Q99PD1;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA nucleotidyltransferase (EC 2.7.7.31) (Terminal addition enzyme)
DE (Terminal deoxynucleotidyltransferase) (TDT) (Terminal transferase).
GN DMT OR TDT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=66286588; PubMed=3755527;
RA Koizumi O., Yokota T., Kageyama T., Hirose T., Yoshida S., Arai K.;
RT "Isolation and characterization of bovine and mouse terminal
RT deoxynucleotidyltransferase cDNAs expressible in mammalian cells.";
RL Nucleic Acids Res. 14:5777-5792(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BAIR/C; TISSUE=Thymus;
RX MEDLINE=93215079; PubMed=8464703;
RA Doyen N., Fanton D'Andon M., Bentolila L.A., Nguyen T.O., Rougeon F.;
RT "Differential splicing in mouse thymus generates two forms of
RT terminal deoxynucleotidyl transferase.";
RL Nucleic Acids Res. 21:1187-1191(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND CHARACTERIZATION.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RX PubMed=11136823;
RA Benedict C.L., Gillilan S., Kearney J.F.;
RT "The long isoform of terminal deoxynucleotidyl transferase (Tdtl)
RT enters the nucleus and, rather than catalyzing N addition, modulates
RT the catalytic activity of the short isoform.";
RL J. Exp. Med. 193:89-99(2001).
RN [4]
RP PRELIMINARY CHARACTERIZATION OF ALTERNATIVE FORMS.
RX MEDLINE=96016194; PubMed=7556063;
RA Bentolila L.A., Fanton D'Andon M., Nguyen T.O., Martinez O.,
RA Rougeon F., Doyen N.;
RT "The two isoforms of mouse terminal deoxynucleotidyl transferase
RT differ in both the ability to add N regions and subcellular

```

RT Localization.
RL EMBL J. 14:4221-4229(1995).
CC -1- FUNCTION: TEMPLATE-INDEPENDENT DNA POLYMERASE WHICH CATALYZES THE
CC RANDOM ADDITION OF DEOXYNUCLEOSIDE 5'-TRIPHOSPHATE TO THE 3' END OF
CC A DNA INITIATOR. ONE OF THE IN-VIVO FUNCTION OF THIS ENZYME IS THE
CC ADDITION OF NUCLEOTIDES AT THE JUNCTION (N REGION) OF REARRANGED
CC HEAVY CHAIN AND T CELL RECEPTOR GENE SEGMENTS DURING THE
CC MATURATION OF B AND T CELLS. The TdtL isoform seems to serve to
CC modulate the function of Tdts.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate +
CC (deoxynucleotide)(M) = N diphosphate + (deoxynucleotide)(M+N).
CC -1- COFACTOR: REDUCES MAGNESIUM.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: Tdt-L(arge)/TdtL (shown here)
CC and Tdt-S(mall)/Tdts, are produced by alternative splicing. The
CC Tdt-S form is the major form.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-X FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BRC1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04123; CAA27735.1; -
CC EMBL: X68670; CAA48634.1; -
CC EMBL: AF316014; AAK07884.1; -
CC EMBL: AF316015; AAK07885.1; -
CC PIR: B23595; B23595.
CC HSSP: P06766; 1BPB.
CC MGD: MGI:98659; DntL.
CC InterPro: IPR001357; BRC1.
CC InterPro: IPR002054; DNA_polX.
CC Pfam: PF00533; BRC1; 1.
CC Pfam: PF00966; DNA_polymeraseX; 1.
CC PRINTS: PR00869; DNAPOLX.
CC SMART: SM00292; BRC1; 1.
CC SMART: SM00483; POLX; 1.
CC PROSITE: PS50172; BRC1; 1.
CC PROSITE: PS50522; DNA_POLYMERASE_X; 1.
CC Transferrase; Nucleotidyltransferase; Terminal addition; Magnesium;
CC Nuclear protein; Alternative splicing.
CC KW Domain 27 118 BRC1.
CC FT VARSPPLIC 483 502 MISSING (IN ISOFORM TDT-S).
CC FT CONFLICT 36 36 T->M (IN REF. 2).
CC FT CONFLICT 99 99 L->F (IN REF. 2).
CC FT CONFLICT 193 193 R->G (IN REF. 1).
CC FT CONFLICT 287 287 Q->K (IN REF. 1).
CC FT CONFLICT 309 309 E->Q (IN REF. 1).
CC FT CONFLICT 367 367 D->H (IN REF. 1).
CC FT CONFLICT 441 444 DRRA->ECAC (IN REF. 1).
CC FT CONFLICT 443 445 RAE->AS (IN REF. 2).
CC FT SEQUENCE 530 AA; 60331 MW; E6B109DCF39C8107 CRC64;
SQ

```

Query Match 6.7%; Score 115; DB 1; Length 530;

Best Local Similarity 24.0%; Pred. No. 0.19; Indels 60; Gaps 16;

Matches 71; Conservative 49; Mismatches 116;

```

DB 8 DLEETPLDDDLTTEFKILAYTRHHVFKSTPALFSPKLTSTRSLASRGNCNSANESW 67
DB 170 DALDLAENDELKENSCLAFPMASVLSLPPPI-----TSKKDEGI-PCJGDYKX 222
DB 68 TEVSWPCRNSSSE-KAINLGKKSSWKAFA--FGVEKEDSQSTPAVSAQGRLL-EY 122
DB 223 SLIEGIEGSESEAKAVLNDERKSKFLFVSYGVLK-----TAEKWFMRGFTLSKI 277
DB 123 QDSHSGQMSRC-----LSNVEQCLEHENDPVIISIANVAELIYWS-PPQATQAG 174
DB 278 QSDKSLRFQOMQKAGFLYEDLVSCVNR---PEEAVALVKEAVVTFPLDALTMTGCG 333

```

```

OY 175 FKSKEIFVTEGLSFOLOGH-----VPVASSKKDEEOLIAKIVELLYKSG-----DQLER 225
OY 334 FRRC-----KMGHDVDFLTSPEDATEDEEQQLHKHTDFMKQOGLLYLCILES 383
OY 226 KKKDKA-----LMGHFQDGLSYVEFTTQDVLMGVDPRESEVKAQGFALV 275
OY 384 TFEFKOPSRKVDALDHFOK--CFILK--LDHGRVHSEKSGQEGK--GMAIRV 433
DB

```

RESULT 6

ID	BC1X_PIG	STANDARD;	PRT;	233 AA.
AC	077737;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Apoptosis regulator Bcl-x.			
GN	BCL2L1 OR BCL2L OR BCLX.			
OS	Sus scrofa (Pig).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Bartling B., Hoffmann J., Holtz J., Schulz R., Heusch G., Darmer D.;			
RT	Expression of apoptosis-associated genes in hibernating and stunned			
RT	myocardium of pig ";			
CC	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: Potent inhibitor of cell death. Inhibits activation of			
CC	caspases (By similarity). Appears to regulate cell death by			
CC	blocking the voltage-dependent anion channel (VDAC) by binding			
CC	to it and preventing the release of the caspase activator,			
CC	cytochrome c, from the mitochondrial membrane.			
CC	-1- SUBUNIT: Bcl-x(L) forms heterodimers with BAX, BAK and Bcl-2 (By			
CC	similarity). Heterodimerization with BAX does not seem to be			
CC	required for anti-apoptotic activity (By similarity).			
CC	-1- SUBCELLULAR LOCATION: MITOCHONDRIAL MEMBRANES AND PERINUCLEAR			
CC	ENVELOPE (BY SIMILARITY).			
CC	-1- DOMAIN: The BH4 domain is required for both heterodimerization			
CC	with other Bcl2 family members and for repression of cell death.			
CC	-1- PTM: Proteolytically cleaved by caspases during apoptosis (By			
CC	similarity). The cleaved protein, lacking the BH4 domain, has pro-			
CC	apoptotic activity (By similarity).			
CC	-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 1 (BH1).			
CC	-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 2 (BH2).			
CC	-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 3 (BH3).			
CC	-1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 4 (BH4).			
CC	-1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: AJ001203; CAA04597.1; -			
DR	HSSP: Q07817; IMAZ.			
DR	InterPro: IPR002475; BCL2_family.			
DR	InterPro: IPR003093; BH4.			
DR	InterPro: IPR000712; BCL_2.			
DR	Pfam: PF00452; Bcl-2; 1.			
DR	Pfam: PF02180; BH4; 1.			
DR	SMART: SM00357; BCL; 1.			
DR	SMART: SM00265; BH4; 1.			
DR	PROSITE: PS50062; BCL2_FAMILY; 1.			
DR	PROSITE: PS01080; BCL; 1.			
DR	PROSITE: PS01258; BH2; 1.			
DR	PROSITE: PS01259; BH3; 1.			
DR	PROSITE: PS01260; BH4_1; 1.			

reticulum.";
 RL J. Biol. Chem. 270:9327-9333(1995).
 CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE A
 CC AND ANCHORS/TARGETS THEM TO THE NUCLEAR MEMBRANE OR SARCOPLASMIC
 CC RETICULUM. MAY ACT AS AN ADAPTER FOR ASSEMBLING MULTIPROTEIN
 CC COMPLEXES.
 CC -1- SUBUNIT: INTERACTS WITH RII SUBUNIT OF PKA. PHOSPHATASE 2B
 CC (CALCIUM-BINDING) AND AKAP79.
 CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM AND NUCLEAR MEMBRANE
 CC IN HEART MUSCLE. PARTICIPATION OF MULTIPLE TARGETING SIGNALS ALLOW
 CC CORRECT INTRACELLULAR TARGETING. THESE MAY BE REPEATED MOTIFS RICH
 CC IN BASIC AND HYDROPHOBIC AMINO ACIDS, PALMITOYLATED/MYRISTOYLATED
 CC MOTIFS OR ALTERNATIVELY SPICE TARGETING SEQUENCES.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN CARDIAC AND SKELETAL
 CC MUSCLE, FOLLOWED BY BRAIN.
 CC -1- DOMAIN: RII-ALPHA BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC
 CC HELIX, COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
 CC -1- SIMILARITY: CONTAINS 2 SPECTRIN REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U17195; AAA92354.2; -
 CC EMBL: AB002309; BAA20770.1; -
 CC MIM: 604691; -
 CC InterPro: IPR002017; Spectrin.
 CC DR Pfam: PF00435; spectrin; 2.
 CC DR SMART: SM00150; SPEC; 2.
 CC KW Repeat.
 CC FT REPEAT 762 848 SPECTRIN 1.
 CC FT REPEAT 1036 1150 SPECTRIN 2.
 CC FT DOMAIN 1560 1701 SER-RICH.
 CC FT DOMAIN 2063 2076 PKA-RII SUBUNIT BINDING DOMAIN.
 CC FT CONFLICT 974 974 C -> W (IN REF. 2).
 CC FT CONFLICT 1492 1492 V -> A (IN REF. 2).
 CC SEQUENCE 2319 AA; 256663 MW; 3D3037AE40FCFEE CRC64;
 SO
 Query Match 5.6%; Score 96; DB 1; Length 2319;
 Best Local Similarity 20.5%; Pred. No. 37;
 Matches 47; Conservative 44; Mismatches 86; Indels 52; Gaps 9;
 Db 77 S0SSEKAINLGK-----KSSWKAFFGVVERKEDSSTPAKVSAGQRTLEYOD 124
 699 SSSSDIASISGESIESGPLSDILSDESSMPLGKMKRYADEKSERASS-----EKNE 751
 QY 125 SHSQQMSRCSNVEQCLEH-----EAVDPKVISIANRVAEIV-----SWPPQAT 170
 Db 752 SHSATRSALIQKIMODIQHDNYEAIWEKIGFVWKIDEFQIMLEAMETTNWTPPKAE 811
 QY 171 QAGGFSKEIFVEGSLFQLOGHVPAVSSSKKDEEQIILAKI-----ELTKSGQD 222
 Db 812 ----MDIKIKYLETHTLSFKLVNDSHCALKEAVEBEGHOLLEILASHKAGLKMRLRAISQ 867
 QY 223 ---LEKRLKDKALMGHFQDGLSYVFET---ITDQVLMGVDPRESEV 265
 Db 868 WKELQRIKHQHSWILRALDITKAEILATDVSVEDEGTG-SPKAEVL 915
 RESULT 11
 BAA_MOUSE STANDARD; PRT; 192 AA.
 AC 007813;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Apoptosis regulator Bax, membrane isoform alpha.

GN BAX.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6 X DBA/2;
 RX MEDLINE=9364978; PubMed=8358790;
 RA Oliva J.N., Millman C.L., Korsmeyer S.J.;
 RT "Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that
 RT accelerates programmed cell death.";
 RL Cell 74:609-619(1993).
 CC -1- FUNCTION: ACCELERATES PROGRAMMED CELL DEATH BY BINDING TO, AND
 CC ANTAGONIZING THE APOPTOSIS REPRESSOR BCL-2 OR ITS ADENOVIRUS
 CC HOMOLOG BCL-19K PROTEIN, INDUCES THE RELEASE OF CYTOCHROME C,
 CC ACTIVATION OF CASPASE-3, AND THEREBY APOPTOSIS. BAX DEFICIENCY
 CC LEADS TO LYMPHOID HYPERPLASIA AND MALE STERILITY, BECAUSE OF THE
 CC CESSATION OF SPERM PRODUCTION.
 CC -1- SUBUNIT: FORMS HOMODIMERS AND HETERODIMERS TOGETHER WITH BCL-2,
 CC E1B 19K PROTEIN, BCL-X(L), MCL-1 AND A1.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.
 CC -1- ALTERNATIVE PRODUCTS: A 21 kDa MEMBRANE PROTEIN ALPHA AND THE TWO
 CC CYTOPLASMIC PROTEINS BETA AND GAMMA ARE GENERATED BY ALTERNATIVE
 CC SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.
 CC -1- DOMAIN: INTACT BH3 DOMAIN IS REQUIRED BY BIK, BID, BAX, BAD AND
 CC BAX FOR THEIR PRO-APOPTOTIC ACTIVITY AND FOR THEIR INTERACTION
 CC WITH ANTI-APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.
 CC APOPTOTIC MEMBERS OF THE BCL-2 FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 1 (BH1).
 CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 2 (BH2).
 CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 3 (BH3).
 CC -1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L22472; AAA03622.1; -
 CC DR HSP: 007817; IMA2.
 CC DR MGI: 99702; Bax.
 CC DR InterPro: IPR002475; BCL2_family.
 CC DR Pfam: PF00452; Bcl-2; 1.
 CC DR SMART: SM00337; BCL; 1.
 CC DR PROSITE: PS01080; BH1; 1.
 CC DR PROSITE: PS01258; BH2; 1.
 CC DR PROSITE: PS01259; BH3; 1.
 CC DR PROSITE: PS50062; BCL2_FAMILY; 1.
 CC KW Apoptosis; Transmembrane; Alternative splicing.
 CC FT DOMAIN 59 73 BH1.
 CC FT DOMAIN 98 118 BH2.
 CC FT DOMAIN 150 165 BH3.
 CC FT TRANSMEM 172 192 POTENTIAL.
 CC SEQUENCE 192 AA; 21394 MW; D2E0B356579FAFF CRC64;
 SO
 Query Match 5.5%; Score 94.5; DB 1; Length 192;
 Best Local Similarity 23.3%; Pred. No. 19;
 Matches 37; Conservative 28; Mismatches 77; Indels 17; Gaps 5;
 Db 172 AGGFSKEIFVEGSLFQLOGHVPAVSSSKKDEEIQ-----LARIYELKYSGD 221
 Db 10 SGGPTSSQIMKTG-AFLQGFQIDRAGMAGETPELTLEQPODASTKLSLSECLRRIGD 68
 QY 222 QLEKRLKDKALMGHFQDGLSYVFETITDQVLMGVDPRESEVKKAGFAALYID-VTA 280
 Db 69 ELDSNMEIQRMITADVDTDSPREVFFRVADMFADGNFMGR-VVALFFASKIVLKALCT 127

QY 261 KLFADNHPNRYLGFCTKYLKENSFPWIOHGWEKIL 319
ID 128 KVEL-IRITMGWTLDFRLRLVWIDOGGMEGL 162

RESULT 12
ARL1_XENLA STANDARD; PRT; 204 AA.
AC 091828;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Apoptosis regulator R11 (XR11).
DE Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-Head:
RA MEDLINE=9531613; PubMed=7607538;
RA Cruz-Reyes J., Tata J.R.,
RA "Cloning, characterization and expression of two Xenopus bcl-2-like
RA cell-survival genes".
RL Gene 158:171-179(1995).
CC -1- FUNCTION: CONFERS STRONG PROTECTION AGAINST CELL DEATH.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -1- DEVELOPMENTAL STAGE: DEVELOPMENTAL REGULATION ONLY OCCURS IN THE
CC BRAIN OF MID-METAMORPHOSIC TO POST-METAMORPHOSIC TADPOLES AND
CC ADULTS, WHERE AN INCREASE OF SEVERAL FOLD HAS BEEN OBSERVED.
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 1 (BH1).
CC -1- SIMILARITY: CONTAINS 1 BCL-2 HOMOLOG DOMAIN 2 (BH2).
CC -1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: X82461; CAAS7844.1; -
DR HSSP: Q07817; 1MAZ.
DR InterPro: IPR002475; BCL2_family.
DR InterPro: IPR003093; BH4.
DR InterPro: IPR000712; BCL_2.
DR Pfam: PF00452; Bcl-2; 1.
DR Pfam: PF02180; BH4; 1.
DR SMART: SM00337; BCL; 1.
DR SMART: SM00265; BH4; 1.
DR PROSITE: PS01080; BH1; 1.
DR PROSITE: PS01258; BH2; 1.
DR PROSITE: PS30062; BCL2_FAMILY; 1.
DR Apoptosis; Transmembrane.
FT DOMAIN 101 120 BH1.
FT DOMAIN 152 167 BH2.
FT TRANSMEM 181 198 POTENTIAL.
SQ SEQUENCE 204 AA; 23379 MW; 3BFC6BEDDA4CA03 CRC64;

Query Match 5.4%; Score 93; DB 1; Length 204;
Best Local Similarity 20.6%; Pred. No. 2,7;
Matches 43; Conservative 34; Mismatches 58; Indels 74; Gaps 9;

QY 131 SACLSEVOCLEHVAQPKYISIANRYAEIVYSPPOATGAGGFKSEIVTEGLISFOL 190
ID 14 SKRLSQNEACR- FSNPNPMYLMETSESRP-EGAT- 50
DB 191 QGHVPAVASSKDEEQLIAKIYVELLKYSQDLERKLKKDAIMGHFDGLSYSEFKTT 250

DB 51 QGIV-EEVLAQILLEATE-EEELRYQRA-FSDLT 81

QY 251 DQVLGADVPGESEVYKAG- FKAALVIDYAK-LTAIDNHPNR 292
DB 82 SGLHITQDTAQSFGQVVGELFRODGTNMGRIYAFPSFGALCVESANKENTDL- LPR 137

QY 293 VLGFCTKYLKENSFPWIOHGWEKIL 321
DB 138 IVQMMVNYLEHTLQPMQENGMEAFVGL 166

RESULT 13
PMS2_MOUSE STANDARD; PRT; 859 AA.
ID PMS2_MOUSE
AC P54279;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE PMS1 protein homolog 2 (DNA mismatch repair protein PMS2).
GN PMS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=95354212; PubMed=7628019;
RA Baker S.M., Bronner C.E., Zhang L., Plug A., Robatzek M.,
RA Warren G., Elliot F.A., Yu J., Ashley T., Anheilm N.,
RA Flavell R.A., Liskay R.M.,
RT Male mice defective in the DNA mismatch repair gene PMS2 exhibit
RT abnormal chromosome synapsis in meiosis.";
RL Cell 82:309-319(1995).
CC -1- FUNCTION: PROBABLY INVOLVED IN THE REPAIR OF MISMATCHES IN DNA.
CC -1- SUBUNIT: HETERODIMER OF PMS2 AND MSH1.
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U28724; AAA87031.1; -
DR HSSP: P23367; 1BKN.
DR MGD: MGI:104288; Pms2.
DR InterPro: IPR002099; DNA_mis_repair.
DR InterPro: IPR003594; HATPase_C.
DR Pfam: PF01119; DNA_mis_repair; 1.
DR Pfam: PF02518; HATPase_C; 1.
DR PROSITE: PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW DNA repair.
SQ SEQUENCE 859 AA; 95225 MW; 263B5A6BBB2AC9 CRC64;

Query Match 5.4%; Score 93; DB 1; Length 859;
Best Local Similarity 22.2%; Pred. No. 17;
Matches 61; Conservative 46; Mismatches 124; Indels 44; Gaps 10;

QY 53 LSQRL-GNCSANESWTEVSWPCRNSSSEKAINIGKKSSWKAFFGVYEKEDSOSTPAK 111
DB 463 ISYGLRGSQDKIVSPDSDPCMDREKIEKDSGLSTAGSEEFSTPEVASSFSDDYN 522

QY 112 VSAGQGRLERFQDSHSQWMSR-LSNVEQL-EHEANDPVYISIANRYAEIVYS 163
DB 523 VSS-LEDPSQETINCGLDRCPPGTGQSLKPPDHGYQCKALPLARL- 568

QY 164 WPPQATGAGGFKSEIVTEGLSFOLQGHVPAVASSKDEEQLIAKIYVELLKYSQDL 223
DB 569 -SPTNAKFKTEERFSNVNISOQLRG-POSTSAALVDAIKMKRIVLLEFSLSSL 622

OY 224 ERKLRKDKALMGHFQDGLSYSVFKTTIDVLMGVDPGRSESEYKAGCFKALY-IDVTAKL 282
 DB 623 AKRMKQLOHLKAKNKHLSYKFRFA---KICPGNOADELKRELSKMFAMEILGOF 679
 OY 283 TAIIDNHPMNRVLGFGTGYKLENFSPWIOGHGWK 317
 DB 680 N-----LGFIVTKLEKEDLF-LVDQHADER 703
 RESULT 14
 GGAS_HUMAN STANDARD; PRT: 723 AA.
 ID GGA3_HUMAN Q9N552; Q9JUY3; Q15017;
 AC 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ADP-ribosylation factor binding protein GGA3 (Golgi-localized, gamma
 ear-containing, Arf-binding protein 3).
 DE GGA3 OR KIA0154.
 DE Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
 RC TISSUE=Heart;
 RX MEDLINE=20211638; PubMed=10747089;
 RA Dell'Angelica E.C., Puertollano R., Mullins C., Aguilar R.C.,
 RA Vargas J.D., Hartnell L.M., Bonifacio J.S.;
 RT "Ggas" a family of Adp-ribosylation factor-binding proteins related to
 RT adaptors and associated with the Golgi complex.";
 RL J. Cell Biol. 149:81-94(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=20214818; PubMed=10749927;
 RA Boman A.L., Zhang C.-J., Zhu X., Kahn R.A.;
 RT "A family of Adp-ribosylation factor effectors that can alter
 RT transport through the trans-Golgi."
 RL Mol. Biol. Cell 11:1241-1255(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RC TISSUE=Bone marrow;
 RX MEDLINE=96127530; PubMed=8590280;
 RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. IV.
 RT The coding sequences of 40 new genes (K1A0121-K1A0160) deduced by
 RT analysis of cDNA clones from human cell line KG-1."
 RL DNA Res. 2:167-174(1995).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE REGULATION OF MEMBRANE TRAFFIC
 CC THROUGH THE TRANS-GOLGI NETWORK.
 CC -1- SUBUNIT: BINDS TO ARF1.
 CC -1- SUBCELLULAR LOCATION: TRANS-GOLGI NETWORK.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: UBIDITUOUSLY EXPRESSED.
 CC -1- SIMILARITY: CONTAINS 1 GAMMA-ADAPTIN C-TERMINAL DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 VHS DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AF219138; AAF42848.1; -
 DR EMBL: AF190864; AAF05709.1; -
 DR EMBL: AF219139; AAF42849.1; -
 DR EMBL: D63876; BAA09926.1; ALT_INIT.
 DR InterPro: IPR001121; G_adapt_C.
 DR InterPro: IPR002014; HRS.
 DR Pfam: PF00790; VHS; 1.

DR Prodom; PD003686; HRS; 1.
 DR Prodom; PD021457; G_adapt_C; 1.
 DR SMART: SMO0288; VHS; 1.
 DR PROSITE; PS50179; VHS; 1.
 DR Protein transport; Golgi stack; Coiled coil; Alternative splicing.
 KW DOMAIN 1 313 BINDS TO ARF1 (IN LONG ISOFORM).
 FT DOMAIN 16 146 VHS.
 FT DOMAIN 188 232 COILED COIL (POTENTIAL).
 FT DOMAIN 357 360 POLY-PRO.
 FT DOMAIN 453 457 POLY-SER.
 FT DOMAIN 598 709 POLY-VAL.
 FT VARSPIC 68 100 GAMMA-ADAPTIN C-TERMINAL.
 FT SEQUENCE 723 AA; 78314 MW; 4F80D6032239168C CRC64;
 Query Match 5.4%; Score 92.5; DB 1; Length 723;
 Best Local Similarity 18.4%; Pred. No. 15;
 Matches 65; Conservative 49; Mismatches 85; Indels 155; Gaps 17;
 OY 65 ESWTEVSPPCRNSSQSEKAINLQKKSSKAFPGVYEKEDSOSTPAKVSAGQRTLEYOD 124
 DB 10 ESWL-----NKAATPSNRQEDWEYIIIGFCDQINKELEGQIAV--RLLAHRI 54
 OY 125 SHSQOWS--RCLSNVEQCLEH-----EAVDPKYI-----SIANRV 157
 DB 55 QSPQEWELQALITVEACMKNGCRPHNEVGKRFLELIRKIVSPKILGDRVSEKATKY 114
 OY 158 AEIVYSW-----PPQATQAGGFSKE 179
 DB 115 IELYSWTMALPEEAKIKDAYHMLKRGQIVQSDPPIPVDRTLIPSPPR-----PKNP 167
 OY 180 IFVTEGLSFQ-----QGHVYASSKKDE-----EQ-----QI 208
 DB 168 VFDEEKSLLAKLIKSNKNDPLQENANKLIKSNVKKDEARIQVTKLHLTEVNNVRL 227
 OY 209 LAKIV-----ELKYSQDLERKLK-----KDKALMGHFQ--DGLS 242
 DB 228 LSEMLIHYSOEDSSDGDRELMKEIPDCENKRTTKLASSETDNDNSLDIIQASNLS 287
 OY 243 YSV--FKTTD-QVLMG-----VDPGRSEYKAGCFKALY-IDVTAKLTAI 285
 DB 288 RVINSYKTIIEGVINGEATLTPDSEGNQSCNCGTILDLIELDPTNTSLSSV 341
 RESULT 15
 RPN2_YEAST STANDARD; PRT: 945 AA.
 AC P32565;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 26S proteasome regulatory subunit RPN2.
 GN RPN2 OR SEN3 OR YIL075C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96026012; PubMed=7565784;
 RA Demarini D.J., Papa F.R., Swaminathan S., Ursic D., Rasmussen T.P.,
 RA Culbertson M.R., Hochstrasser M.;
 RT "The yeast SEN3 gene encodes a regulatory subunit of the 26S
 RT proteasome complex required for ubiquitin-dependent protein
 RT degradation in vivo."
 RL Mol. Cell. Biol. 15:6311-6321(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Cossey T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,

[illegible]

Search completed: June 19, 2002, 16:40:06
Job time: 379 sec

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-562 <BEV>
A:Cross-references: EMBL:AL351013; GSPDB:GN00063; ATSP:T24H18.70
C:Experimental source: cultivar Columbia; BAC clone T24H18
C:Genetics:
A:Gene: ATSP:T24H18.70
A:Map position: 5
A:Introns: 73/3

Query Match 6.7%; Score 115; DB 2; Length 562;
Best Local Similarity 21.7%; Pred. No. 0.4;
Matches 80; Conservative 43; Mismatches 136; Indels 110; Gaps 16;

31 TRHHFKSPALFSPKL-----LRTSL-----ORGLGCSANESWTEVS 71
128 SHHHYTMDSALRSPPLHFTTGRSGSVDFRSVSCNDYKKGKGFDTKLSMLV- 166
72 WPCRNQS--SEKAINLG-----KKSSWKAFF--GVVEKDSQSTPAKVA 114
187 -PLTDSHSAVVSOPNRNGRVMWMLFPKTKKOKNSIFNPSITEKSEVEVLKDSG 245
115 OGQRLENG-----DSHSQMSRCLSNVEQLEH---FAVDPKVISIANRVAEIV 162
246 SGVEKLKRLMEANRSRDALTYVSEKSSIGELSEKLQYLESYCDNLKALREATEV- 304
163 SMPPOATGAG--GPKSKEIFYTEGLSFOGHVPAVSS-----SKKDEE 206
305 ----SQENSGGSSGKKNSMPYSE--EVWVEGFLQIVSEARLSIKQFLKTVSEIDEED 358
207 QILAKIVELLKYSQDOLERRLKDKALMGHFQDGLSYVFKTITDOVLG-----VDPR 260
359 STLLININLT-LQPHNLSEFSKYSKIQYHLEALISQSVYODFEKCVFOKNGKPLDPE 417
261 GSEYKAGCFKALVADVTKLTAIDNHPRNRYLGFCTKYLKENS----- 306
418 QDRQANFSSF-----ASLRNLSMNEVLKKGKYYSDERSFCDEKMSLITTL 465
307 ---PMIOQ 311
466 WMTKRWSEQ 474

RESULT 9
T94912
hypothetical protein BID4.270 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
Accession: T94912
hulte, U.; Align, V.; Hohelsel, J.; Brandt, P.; Farmann, B.; Holland, R.; Nyakatura,
Submitted to the Protein Sequence Database, May 2000
A:Reference number: 225022
A:Accession: T94912
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1647 <SCH>
A:Cross-references: EMBL:AL355928; GSPDB:GN00116; NCSP:BID4.270
C:Experimental source: BAC clone BID4; strain OR74A
C:Genetics:
A:Gene: NCSP:BID4.270
A:Map position: 6

Query Match 6.6%; Score 113.5; DB 2; Length 1647;
Best Local Similarity 23.2%; Pred. No. 2.3;
Matches 71; Conservative 46; Mismatches 102; Indels 87; Gaps 15;

77 SOSSEKAINLG-----KKSSWKAFFGVVEKEDSQSTPAKVASAQO----- 117
422 SPDSAKVELGQDHVGEGRKESVAEAVVGLAEQADREAQELARAEQSAQAGITQDLE 481
118 --RTLEYODSHSQMSRCLSNVEQCLEHE-----AVDPKVISIANRVAEIVSWPPP 167

DB 482 QMSAETDDEFARSQTAQAIOKOLEKDNKVLESALDAV---AKEVRSIV-----D 533
168 QATQAG-----GPKSKEIFYTEG-----LSFOLGCHVPAVSSSKKDE-----E 205
534 KAAQAGCGPVAWDSWESADEIVIEEAPRIKYVNFPLKPMITTI-SVETEDRVRYFRE 592
206 EQILAKIVELLKYSQDOLERRLKDKALMGHFQDGLSYVFKTITDOVLG---GVDPGE 262
593 EALID--IARLKKDFQIDRN-----LVSASESYMAVGMASKAGGLRVIOEDGDKAKLF 644
263 SEYKAGCFKALVADVTKLTAIDNHPRNRYLGFCTKYLKENSFPVIOHGNKEL 319
645 TPTKRIFFNVAI-----SSASNQHPKEAIIIGVSGIVY-----WVQLKNGDRDL 691
320 GISHEE 325
692 EDAHPE 697

RESULT 10
B23595
DNA nucleotidylexotransferase (EC 2.7.7.31) - mouse
N:Alternate names: terminal addition enzyme; terminal deoxynucleotidyltransferase
C:Species: Mus musculus (house mouse)
C:Date: 02-Jun-1988 #sequence_revision 02-Jun-1988 #text_change 18-Jun-1999
Accession: B23595; S30235
R:Koiwai, O.; Yokota, T.; Kageyama, T.; Hirose, T.; Yoshida, S.; Arai, K.
Nucleic Acids Res. 14, 5777-5792, 1986
A:Title: Isolation and characterization of bovine and mouse terminal deoxynucleotidyl
A:Reference number: A93633; MUID:86286588
A:Accession: B23595
A:Molecule type: mRNA
A:Residues: 1-529 <KNO>
A:Cross-references: GB:X04123; NID:954765; PIDN:CAA27735.1; PID:954766
R:Dojyen, N.; d'Andon, M.F.; Bentolila, L.A.; Nguyen, Q.T.; Rougeon, F.
Nucleic Acids Res. 21, 1187-1191, 1993
A:Title: Differential splicing in mouse thymus generates two forms of terminal deoxyn
A:Reference number: S30235; MUID:93219079
A:Accession: S30235
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-25 'M', 27-98 'F', 100-192 'R', 194-286 'Q', 288-308 'E', 310-366 'D', 368-44
A:Cross-references: EMBL:X68670; NID:9287808; PIDN:CAA48634.1; PID:9287809
C:Superfamily: DNA nucleotidylexotransferase
C:Keywords: alternative splicing; magnesium; nucleotidyltransferase; nucleus

Query Match 6.6%; Score 113; DB 2; Length 529;
Best Local Similarity 24.0%; Pred. No. 0.53;
Matches 71; Conservative 48; Mismatches 117; Indels 60; Gaps 16;

8 DLEETPLDDDDLTIEKILLAYTRHHFKSPALFSPKRLTRLSQGLGCSANESW 67
170 DAIDLTAENDELKRENSGCLAIFMGASVYLKSLPPT-----TSKMDTGT-PCLDKVK 222
68 TEVSWPCRNQSSE-KAINEGKKSSWKAFF--GVVEKEDSQSTPAKVASAQOQLT-EX 122
223 STIEGIIIEGSESEAKVANDERKSFRLFSVREVGK-----TAEKFRNGFRRLSKI 277
123 QDSHSQMSRCL-----LSNVEQCLEHEAVDPKYVSIANRVAEIVYSN-PPQATQAG 174
278 QSDKSLRFTKMQKAGFLYEDIVSCVNR---PEAAVSMIVKEAVVTEPLPALVTMTGG 333
175 FSKSEIFVTEGLSFOGLH---PVASSSKKDEEQILAKIVELLKYG-----QDLR 225
334 FRKG-----KMGCHVDFLITSPEATEDEEQDLHKYTHFWKQOGLLYCYILES 383
226 KKKKDKA-----LMGHFQDGLSYVFKTITDOVLGVDPRGESEYKAGCFKALV 275
384 TEKFKQPSKRVADLHFQK--CFLLIK--LDHGIVHSEKSSQOGECK--GMKAIKV 433

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 04:19:16 ; Search time 2402.36 Seconds
(without alignments)
4264.219 Million cell updates/sec

Title: US-09-771-961-3

Perfect score: 759

Sequence: 1 atgtgtagcaccagctggtg.....gttgatgctcattga 759

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estdb:*
2: em_esthm:*
3: em_estln:*
4: em_estm:*
5: em_estor:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_estz:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	637.8	84.0	980	10	BI833102 603090814
2	603.4	79.5	763	10	BG718587 602696672
3	549	72.4	774	10	BG720221 602696672
4	506.6	66.7	715	10	BI838169 603083677
5	497.2	65.5	672	10	BG717835 602693855
6	418.4	55.1	703	10	BI908814 603066277
7	356.8	47.0	422	12	AO131601 HS_3049_A
8	321.4	42.3	406	9	AA399486 zt60c07.r
9	294.6	38.8	1535	11	AK016997 Mus muscu
10	294.6	38.8	1541	11	AK008682 Mus muscu
11	294.6	38.8	1578	11	AK018579 Mus muscu
12	294.6	38.8	2222	11	AK016670 Mus muscu
13	276.2	36.4	776	10	BI463753 603203445
14	246.6	32.5	920	10	BI453888 603174765
15	246.4	32.5	652	9	AA536718 vj88a02.r
16	243.6	32.1	680	10	BG245730 602358519
17	242.6	32.0	448	10	BF198484 248427 NA

18	234.4	30.9	697	10	BF584699 602098438
19	227.2	29.9	924	10	BG176106 602337937
20	209	27.5	666	9	BB628327 BB628327
21	203.8	26.9	1061	10	BF139422 601785402
22	201.2	26.5	403	10	BI342342 370081 MA
23	197	26.0	567	9	AM631581 90021 MAR
24	190.6	25.1	922	10	BB615307 BB615307
25	186	24.5	465	10	BF149673 uy67901.Y
26	149	19.6	423	10	BG895661 359228 MA
27	133.6	17.6	437	9	AA061006 m146f07.r
28	130.6	17.2	697	9	AT614194 v177e07.Y
29	126.6	16.7	676	10	BI652080 60329866
30	123	16.2	354	9	AM414655 48366 MAR
31	110.2	14.5	984	9	BB013694 BB013694
32	109.2	14.4	376	9	AM484860 62957 MAR
33	104.8	13.8	370	9	AM428115 64957 MAR
34	97.6	12.9	216	10	BI360692 388694 MA
35	93.8	12.4	332	10	T69681 yca4f12.r1
36	89.4	11.8	473	9	AA400686 zu70c11.r
37	88.4	11.6	446	9	BB746476 BB746476
38	87.4	11.5	256	9	AI606692 m157b10.Y
39	84.4	11.1	426	9	AM000827 w45d10.x
40	64.6	8.5	195	9	AA107133 m157b10.r
41	59.4	7.8	451	10	BM433532 10m16402
42	57	7.5	269	9	BB576539 BB576539
43	50.2	6.6	514	12	AO884233 HS_5504_B
44	48.4	6.4	184	9	AA492696 v177e07.r
45	40.2	5.3	221	9	BB589687 BB589687

ALIGNMENTS

RESULT 1
BI833102 980 bp mRNA linear EST 04-OCT-2001
LOCUS 603090814F1 NIH_MGC_120 Homo sapiens CDNA clone IMAGE:5229819 5',
DEFINITION mRNA sequence.
ACCESSION BI833102.1 GI:15944652
VERSION BI833102
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgs.ncl.nih.gov/.
1 (bases 1 to 980)
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC Clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: LMN1578 row: k column: 04
High quality sequence stop: 803.
Location/Qualifiers
1. 980
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5229819"
/clone_lib="NIH_MGC_120"
/lab_host="DH10B"
/note="Organ: pooled pancreas and spleen; Vector:
PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 yo
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is

FEATURES

source

	Db	625	GTTTATTCCTGGCCACCAACAAGGACCACGAGGAGGACTTCAAAGTCCAAGAAGATT	684
Oy	541	tttgtaactgtagggctctctcttcacgtcccaagccaagtgcctgtagcttcaagttct	600	
Db	685	TTTGTAACTGAAGGCTCTCTCCTTCACGCTCAAAGGCACAGCTGCTTAGCTTCAAGATTCT	744	
Oy	601	aagaa 605		
Db	745	AAGAA 749		
RESULT	3			
BG720221		774 bp	mRNA	linear
LOCUS		602692271F1 NIH_MGC_97	Homo sapiens cDNA clone IMAGE:4824760 5'	EST 08-MAY--2001
DEFINITION				
ACCESSION		BG720221		
VERSION		BG720221.1	GI:13999408	
KEYWORDS		EST.		
SOURCE		human.		
ORGANISM		Homo sapiens		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
		NIH-MGC http://mgc.nci.nih.gov/.		
		1 (bases 1 to 774)		
REFERENCE		National Institutes of Health, Mammalian Gene Collection (MGC)		
AUTHORS		Contact: Robert Strausberg, Ph.D.		
TITLE		Email: cgabs-remail.nih.gov		
JOURNAL		Tissue Procurement: Miklos Palkovits, M.D., Ph.D.		
COMMENT		cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki		
		Toshiyuki and Piero Carninci (RIKEN)		
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
		DNA Sequencing by: Incyte Genomics, Inc.		
		Clone distribution: MGC clone distribution information can be		
		found through the I.M.A.G.E. Consortium/LNL at:		
		http://image.lnl.gov		
		Place: LHAM10736 row: m column: 17		
		High quality sequence stop: 769.		
FEATURES		Location/Qualifiers		
source		1..774		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone_image="4824760"		
		/clone_idb="NIH_MGC_97"		
		/lab_host="DH10B"		
		/note="Organ: testis; Vector: pBluescriptR (modified		
		pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag		
) ; Oligo-df primed using primer 5'-TTTTTTTTTTTTTTTVN-3' ,		
		size-selected for average insert size 2.2 kb and		
		normalized to ROT 5. This is a primary library enriched		
		for full-length clones and constructed using the		
		Cap-trapper method (Carninci, in preparation). Library		
		constructed by M. Brownstein (NHGRI/NHGRI, National		
		Institutes of Health). Note: this is a NIH-MGC Library."		
BASE COUNT		225 a 185 c 198 g 166 t		
ORIGIN				
Query Match		72.3%; Score 549; DB 10; Length 774;		
Best Local Similarity		98.1%; Pred. No. 1.7e-148;		
Matches 609; Conservative		0; Mismatches 5; Indels 7; Gaps 5;		
Oy	1	atctgtagaaccaggctgtgtacctgtgaanaaacccccctgatgatgatgaacctaac 60		
Db	159	ATGTGTAGCACCGTGAGTGTGACCTGGAAAGAAATCCCCTGATGATGATGACCTTAAC 218		
Oy	61	accatagaattcaaatcctcgtcctactacacccaagacatcatgctctcaagaagcacccct 120		
Db	219	ACCATGGAATTCAAAATCCGCGCTACTAGACACGACATCATGCTTCAAGAACGCCCT 278		
Oy	121	gtctctcttccaccaagctgctgtagaacaagaagtctgtccagaagggcctggygaat 180		

Db 279 GCTCTCTTTCACCAAAGCTGCTGAGAACAAAGAAGTTTGTCCAGAGGGGCTTGGGAAAT 338

QY 181 ttttcaagcaaatgagatcagatcagacagaggtgtcatgttgctctgcagaaattccaatcag 240

Db 339 TTTTATGCAAAATGATGACATGAGACAGAGGTGTATGGCTTTGCAGAAATTTCCCAATTCAG 398

QY 241 gagaagggccataaccttggccaagaaaaagtcctctcttgaaaaatctctttgaagtagt 300

Db 399 GAGAAAGGCATTAACCTTGGCAAGAAAAAGTCTTCTTGAAAGCATTTCTTGGAGTAGTG 458

QY 301 gagaaggaagatctgcagaagcaagcctgcagaaagtcctctgctcagaggtcctaaagagcttg 360

Db 459 GGAAGAGGAAGATTTCCAGAGCAACGGCTGCCAAGGTCTGTGCTCAGGGTCAAGAGACGTTG 518

QY 361 gaattccaagattgcacagccagccagtggtctcaagtgtcttcttctaagtgcagcagctgc 420

Db 519 GAATTCACCAAGATTCGGACACGCCAGCAGctgtgtccaaagctgtcttTTTAACCTGGAGCGATGC 578

QY 421 -ttggagcatgaagctgttgagccccaagaatcttccattgcacaacagtagtctaagt 479

Db 579 TTTGGAGCATGAAAGCTGTGGACCCCAAAATGATTTCCATTGGCAACCGAGTGTAGTAAAT 638

QY 480 tgttatattcctgtgcccaccacccaagaagcgagcccaagcgaggaagctctaagtccaagaagt 539

Db 639 TGTTTACTTCCTGGC--ACACCAACAAGCGA-CCAGGCAAGGAGCTTCAATCCAAAAGAT 695

QY 540 ttttctaactgaggtctctcctcttcagctccaa-ggccaagctgcctgagctctaagt 598

Db 696 TTTTCTAAGTGAAGGCTCTCTCTTCAGGTCCAAAGGCCACGCTGCCTGTAG--TTCAAGTT 753

QY 599 ctaagaagaagatgaagaagaac 619

Db 754 CTAAGACAGATCAAGAAGAAC 774

RESULT	4
LOCUS	B1838169
DEFINITION	B1838169 715 bp mRNA linear EST 04-OCT-2001
ACCESSION	603083677/F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5222875 5',
VERSION	B1838169
KEYWORDS	mRNA sequence.
SOURCE	B1838169.1 GI:15949719
ORGANISM	EST.
REFERENCE	human.
AUTHORS	Homo sapiens
TITLE	Eukaryota, Metazoa: Chordata; Vertebrata; Euteleostomi;
JOURNAL	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
COMMENT	1 (bases 1 to 715)
	NIH-MGC http://mgc.nci.nih.gov/ .
	National Institutes of Health, Mammalian Gene Collection (MGC)
	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Email: cgapsb@remail.nih.gov
	Tissue Procurement: Life Technologies, Inc.
	cDNA Library Preparation: Life Technologies, Inc.
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
	DNA Sequencing by: Incyte Genomics, Inc.
	Clone distribution: MGC clone distribution information can be
	found through the I.M.A.G.E. Consortium/LLNL at:
	http://image.llnl.gov
	Plate: L1AM11560 row: 1 column: 20
	High quality sequence stiers: 715.
FEATURES	location/Qualifiers
source	1..715
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:5222875"
	/clone_1id="NIH_MGC_120"
	/lab_host="DH10B"
	/note="Organ: pooled pancreas and spleen; Vector:
	pcwv-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
	source anonymous pool of spleen and pancreas from 28 yo
	male. Library is oligo-dT primed and directionally cloned

(EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-2.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 025. Note: this is a NIH_MGC Library."

BASE COUNT 212 a 177 c 171 g 155 t
ORIGIN

Query Match 66.7%; Score 506.6; DB 10; Length 715;
Best Local Similarity 99.2%; Pred. No. 3.5e-136;
Matches 509; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 atgttagacacagtggtgtgtgacccggaagaatccccctagatgatgacctaacc 60
|||||
Db 203 ATGTGTAGACACAGTGGGTGTGTGACCTGGAAAGAAATCCCCCTAGATGATGACTTAAC 262
|||||
Oy 61 accatagaattcaaaatcctcgtcctactacacagacatcatgtcttcagaagacccct 120
|||||
Db 263 ACCATAGAAATCAAAATCTCGCTACTACACAGACATCATGTCTTCAGAGCACCCCT 322
|||||
Oy 121 gctcctctcaccacagctgctgagacaagaattgttcccgagagggcctgaggaat 180
|||||
Db 323 GCTCTCTCTCACCACCAAGCTGTGAGAACAGAGTTGTCCAGAGGGCCTGGGGAAT 382
|||||
Oy 181 ttttcagcaaatgagtcagtcagacagagtgatcatgcttcgacgaatcccaatccagt 240
|||||
Db 383 TGTTCAGCAAAATGAGTCATGACAGAGAGTGTCTCATGAGCTTCAGAAATTCACCAAT 442
|||||
Oy 241 gagaagcgcataaaccttgcaagaagaatctctcttggaagaatctcttgagtagtg 300
|||||
Db 443 GAGAAGGCGCATTAACCTTGCGCAAGAAAGTCTCTTGGAAGCATTTCTTGGAATAGTG 502
|||||
Oy 301 gagaaggaagatcgcagagacagcctgccaaggtctctgctcagaggtcgaagagcgttg 360
|||||
Db 503 GAGAAGGAAGATTCGACAGACAGCCTGCGCAAGGTCTCTGCTCAGAGTCAAGAGACGTTG 562
|||||
Oy 361 gaataccaagatcgcacagccagcagtggtccagtgctctctcctaaagtggaagcagtg 420
|||||
Db 563 GAATACCAAGATTCGACAGACAGCAGCACTGGTGCAGGTCTTTTACGTGAGAGAGTGC 622
|||||
Oy 421 ttgtagcatgaagctgttgagaccccaagtcattccattccacacagtagctgaatt 480
|||||
Db 623 TTGGAGCATGAAGACTGTGGAGCCCAAGTCAATTCATTTGCCAACCGAGTAGCTGAATTC 682
|||||
Oy 481 gttattcctgagccacacacacagagccag 513
|||||
Db 683 GTTACTCTCTGACACCAACACAGCAAGCCAG 715
|||||

LOCUS 5 672 bp mRNA linear EST 08-MAY-2001
DEFINITION 60269385F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4826292 5',
mRNA sequence.
ACCESSION BG717835
VERSION BG717835.1 GI:13997022
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 672)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10740 row: m column: 13
High quality sequence stop: 648.
Location/Qualifiers

FEATURES
source

1. 672
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4826292"
/clone_1db="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescript (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgaag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R07 5. This is a primary library enriched for full-length clones and constructed using the Gap-trap method (Carninci, in preparation). Library constructed by M. Brownstein (NHGRI/NHGR1, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 197 a 162 c 171 g 142 t
ORIGIN

Query Match 65.5%; Score 497.2; DB 10; Length 672;
Best Local Similarity 99.2%; Pred. No. 1.8e-133;
Matches 510; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Oy 1 atgtgtgacacagtggtgtgtgacccggaagaatccccctagatgatgacctaacc 60
|||||
Db 159 ATGTGTAGACACAGTGGGTGTGTGACCTGGAAAGAAATCCCCCTAGATGATGACTTAAC 218
|||||
Oy 61 accatagaattcaaaatcctcgtcctactacacagacatcatgtcttcagaagacccct 120
|||||
Db 219 ACCATAGAAATCAAAATCTCGCTACTACACAGACATCATGTCTTCAGAGCACCCCT 278
|||||
Oy 121 gctcctctcaccacagctgctgagacaagaattgttcccgagagggcctgaggaat 180
|||||
Db 279 GCTCTCTCTCACCACCAAGCTGTGAGAACAGAGTTGTCCAGAGGGCCTGGGGAAT 338
|||||
Oy 181 ttttcagcaaatgagtcagtcagacagagtgatcatgcttcgacgaatcccaatccagt 240
|||||
Db 339 TGTTCAGCAAAATGAGTCATGACAGAGAGTGTATGCTCTGCTCAGAGGTCAAGAGACTTGC 398
|||||
Oy 241 gagaagcgcataaaccttgcaagaagaatctctcttggaagaatctcttgagtagtg 300
|||||
Db 399 GAGAAGGCGCATTAACCTTGCGCAAGAAAGTCTTCTGGAAGCATTTCTTGAGTAGTG 458
|||||
Oy 301 gagaaggaagatcgcagagacagcctgcgaaggtctctgctcagaggtcgaagagcgttg 360
|||||
Db 459 GAGAAGGAAGATTCGACAGAGCAGCCTGCGCAAGGTCTCTGCTCAGAGGTCAAGAGACTTGC 518
|||||
Oy 361 gaataccaagatcgcacagccagcagtggtccagtgctctcctaaagtggaagcagtg 419
|||||
Db 519 GAATACCAAGATTCGACAGACAGCAGTGTGTCAGAGTGTCTTCTTACCTGGAAGAGTGC 578
|||||
Oy 420 ctgtagcatgaagctgttgagaccccaagtcattccattccacacagtagctgaatt 479
|||||
Db 579 CTGAGAGCATGAAGCTGTGGAGCCCAAGTCAATTCATTTGCCAACCGAGTAGCTGAAT 638
|||||
Oy 480 ttttattcctgagccacacacacagagccag 513
|||||
Db 639 GTTTATCTCTGTGACACCAACAGCAAGCCAG 672
|||||

RESULT 6
BI908814 703 bp mRNA linear EST 16-OCT-2001
LOCUS BI908814
DEFINITION 60306627F1 NIH_MGC_118 Homo sapiens cDNA clone IMAGE:5215218 5',
mRNA sequence.
ACCESSION BI908814
VERSION BI908814.1 GI:16171889

QY	308	aagattgcgaagaagacgccttcgaaggtccttgctcgaagtcgaagaagacgttgaataac	367
Db	276	AAGATTGCAAGACGACGCTTCGCAAGGCTCTTGCTCCAGTGCAAGGACGATTGGAAATAC	335
QY	368	aagattgcacacagcgaagtcggtcccaagtgctcttccaagtcgtagacgagtcgttgagc	427
Db	336	AAGATTGCAAGACGACGAGTGCTCCAGTGTCTTTCTAACGTGGAGCAGTGTCTTGAGAC	395
QY	428	atgaagctgtgaccccaaatcat	453
Db	396	ATGAAGGTGACGATCTGGGATTCTTT	421

	RESULT	8
AA399486		
LOCUS	406 bp	mRNA
DEFINITION	z160c07.r1 Soares_testis1_NHT Homo sapiens CDNA IMAGE:726732	linear EST 16-MAY-1997
DESCRIPTION	5', mRNA sequence.	

ANIMALISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 406)
AUTHORS	Hillier, L., Allen, M., Bowles, L., Dubouque, T., Geisels, G., Jost, S.,

Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28ml3 rev2^{ET} from Amesham.

FEATURES		SOURCE	
Location/Qualifiers		1. .406	
		/organism="Homo sapiens"	
		/db_xref="GDB:592364.3"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:726732"	
		/clone_1lb="Soares_testis_NHT"	
		/sex="male"	
		/lab_host="DH10B"	
		/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories Inc. and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAATGGGACGGCGGCCGCCCAATTTTTTTTTTTTT 3'].	
		Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT733 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "	
BASE COUNT	127 a	95 c	101 g 82 t 1 others
ORIGIN			

Qy	1	atgtttagaccacagtggtgttgcacttggaagaatcccccatgatgatgatgaac	60
Db	52	ATGTATACACACAGTGGGTGCACCTGGAGAGAAATCCCCCTGATGATGATACCTAAAC	111
Qy	61	acacatagattcaaatctctgcctctacacaccagatcatcgtcttcaagaacacct	120

Db	112	ACCATGAAATTCAAATTCCTCGCTACTACACCGACATCATGTCTTCACAGACACCCT	171
OY	121	gcat-ctctccaccaagcgcctgaaagaaagaatttgcaccagaagggccctgggaa	179
Db	172	GCTTGCTTCTCACCAAGCTGCTGAGACACAGAGTTTGTCCAGAGGCCATGGGAA	231
OY	180	ttgttcagcaaatgagtcacgacagagtgatgagccttgaga-aattccaatcca	238
Db	232	TTGATTCACCAATATAGTCATGAGAGAGTGTCATGCGCTTCGAGAAATATCCCAATCCA	291
OY	239	gtggaagggccataaaccttgcaagaataaagctctctgaaagcatctcttgagtag	298
Db	292	GTGGAAGGCCATTAACCTTGCAAGAAAGTGCTTCTTGAAAGCATCTTGTGAGTAG	351
OY	299	tgaggaaggaagatttcgacagagacgcttcgcaaggtctgcaccaggttcaag	353
Db	352	TGGAGAAAGGAATTCGAGAGCACGCTGCCAAGGTCTTGCTCAGGGTCAAG	406

RESULT	9
AK016997	
LOCUS	1535 bp mRNA
DEFINITION	linear HTC 19-JAN-2002
	Mus musculus adult male testis cDNA, Riken full-length enriched library, clone:4933430101:B2C2.1like apoptosis inhibitor containing protein, full insert sequence.
ACCESSION	AK016997
VERSION	AK016997.1 GI:12856033
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone 11b:RIKEN full-length enriched mouse cDNA library

ORGANISM	Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (sites)
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2 (sites)
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okaaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20498374

REFERENCE
 PUBLISHED
 11042159
 3 (sites)
 Shihbata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, T., Nishi, K., Kitsuaki, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujiwara, K., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, K.,
 Yoneeda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multipillar sequencer
 Genome Res. 10 (11), 1757-1771 (2000)
 JOURNAL
 MEDLINE
 PUBMED
 20530913
 11076861
 4 (sites)
 THE RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)
 5 (bases 1 to 1535)
 Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
 Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T.,
 Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
 Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,

Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sugabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Riken = 10.0 and subtraction to Riken = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATCCAGAGCTCTTTTCTTTTCTT 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.

FEATURES

source Location/Qualifiers

1..1535
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MG1:1894677"
/db_xref="taxon:10090"
/clone="493430L01"
/sex="male"
/tissue="testis"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
249..1235
/note="Bcl2-like apoptosis inhibitor containing protein data source: InterPro, source key: IPR002475, evidence: ISS putative"
/codon_start=1
/protein_id="BAB30545.1"
/db_xref="GI:12856034"
/translation="MCTSYVDLEDIPLEDPPNSIEKILAFYARHVFKNTPAVESPKLSRSLQKALGTWSTDSVSLPCRSPPSEKINISLQKSSWTLPEVAKEEGLSPSPKEIRAGQPGPFPEVERQSGHNMWPSLSEYQLSESVKAFVIANRVAEIVSWPPEVDVHSQSGSKLKEKRESEILYFRGPGDSDQDIISKVELLKFSGDQGREIKDKKALMSRQDGLSYEFKTTDFLDVDRGSEVAVARGKALALIDIAIKTALIDNPMRMGLGRTYLYRFSPWYQNGWEKILGISHEVD"

CDS

BASE COUNT 435 a 384 c 399 g 317 t
ORIGIN

Query Match 38.88; Score 294.6; DB 11; Length 1535;
Best Local Similarity 69.68; Pred. No. 1.9e-74;
Matches 482; Conservative 0; Mismatches 184; Indels 27; Gaps 5;

QY 1 atgtgtgacacacagtggtgacacaggaagaaccccgatgatgatgacccaac 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 249 ATGTGACACACAGTGTGATGACCTGGAAGACATTCCTCGAGATGATGATCCAAAC 308
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 61 accatagaaattcaaatcctgactactaacacagacatcatgcttcaagagaccct 120
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 309 AGCATAGACTTCAAAAATCCTGGCCTTCTTACGCCAGACACACATGCTCTTCAAGACACCCCG 368
QY 121 gctctctctccacaaagctgtcgaagacaagaagttgtcccaagagggcctggaat 180
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 369 GGTGCTTCTTCCGCCAAGGCTCTCCAGACAGAGAGTGTGTCCAGAAAGCCTGGGACT 428
QY 181 tttcagaacaatgatgatcatgatgacagaggtgtcatgtccttgcagaaatcccaatcagt 240
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 429 TGGT---CAACTGATTCCTGGACACAGAGTATTCATTCCTTGCACAGAGTTCCTCCAGC 485
QY 241 gagaagggccataaaccttggcagaagaaagctctcttgcagaagatcttggagtagtg 300
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 486 GAAAAGAAATACGCTTGAGCAAGAAAGAAAGTCTTGGAGAAACCTTCAGAGGTGACC 545
QY 301 gaga---aagaagatttcgacagacagccgcgcaaggtctctgcagaggtcacaagagcg 357
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 546 GAGAGGAGAGAGAGCGCTCGGACCTCCCAAGAGAGATCCGAGCTCAGAGGTCTCAGGCGC 605
QY 358 ttggaataccaagatgc-----cagaagcagaagtggtccagtgcttctct 405
Db 606 CCTTCCCGGTAGAGCGGAGAGAGTGTTCACACACAGCAGCTGGCCAGGCTCTGAGC 665
QY 406 aacgtgagacatgtcttgagacatgaagctgtgagcccaagtcattccattgccaac 465
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 666 AGTGTGAGACACCGCTGAGAGTGAAGTGTGATTCGAAAGGCGCTTTATTCGCAAC 725
QY 466 cgaatgacgaatgttattctctgcacacacacacacagcagcagcagcagcagc 525
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 726 AGAGTGGCTGAATTTTACTCTGTGCGCCACACACAGATGTCATCCACAGCAGGAGAGA 785
QY 526 aagtcacaagagatttgaactgaggtctctctctccttcagctcacaagcagctgct 585
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 786 AGCAAGCTCAAGAGAGGAGCTCGAGATTTCTGACTTCAAGGTTGAAGGAC-----CTT 840
QY 586 gtagcttcaagttctaaagaagaatgaagaagaacaatactagccaataatgttgagctg 645
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 841 GTGACTGTAAAG---AATAAAGATGAGAGACCAAAATTAATACCAAGATTGTGAGCTG 896
QY 646 ctgaatatctcagagatcagttggaagaag 678
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 897 CTGAATTTCTCGGCGGATCAGTTGCGGAGAGAG 929

RESULT 10
AK008682 1541 bp mRNA linear HTC 19-JAN-2002
LOCUS
DEFINITION Mus musculus adult male stomach cDNA, RIKEN full-length enriched library, clone:2210008009:hypothetical protein, full insert sequence.
ACCESSION AK008682
VERSION AK008682.1 GI:12843025
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain:C57BL/6J) adult male stomach cDNA to mRNA, clone:11b:RIKEN full-length enriched mouse cDNA library
clone:2210008009.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
2 (sites)
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)

SOURCE	Mus musculus (strain:C57BL/6J) adult male colon cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library clone:9030625M01.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)
JOURNAL TITLE	99279253
MEDLINE PUBMED	10349636
REFERENCE	2. (sites) Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE	20499374
PUBMED	11042159
REFERENCE	3. (sites) Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,A., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiyagi,K., Fujisake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanagi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)
TITLE	4. (sites) The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
JOURNAL MEDLINE	11076861
PUBMED	20530913
REFERENCE	5. (bases 1 to 1578) Nature 409, 685-690 (2001)
TITLE	Aachui,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Horii,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Nishiazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shimizu,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sobabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamakoshi,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
JOURNAL TITLE	Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gs.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and genome science laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGAGATCCAAAGACCTCTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.2. Second strand cDNA was prepared with the primer

FEATURES	source
adapter of sequence [5', GAGGAGGAGATTTCGACGTTAATTAATTAATCAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after, bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B. Location/Qualifiers	
1..1578	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="MGI:1894816"
	/db_xref="taxon:10090"
	/clone="9030625M01"
	/sex="male"
	/tissue_type="colon"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
	/dev_stage="adult"
216..1202	/note="evidence:NAS hypothetical protein putative"
	/protein_start=1
	/protein_id="BAB31290.1"
	/db_xref="GI:12858358"
	/translation="MCSYSDYDIEDIPLEDDPNSTIEFKIAFYARHAYKNPAPVSS PKSTRSRISOKALGTWSTDSMTQVSIYICRGRSSSEKNTSLGKKSWRLTFVAEKE EGIPSSPKRIEAGCPQGPPEYEROSGFILNHPRLSSVEORLESEYVDSKAIAR VAIVISWPPPDV IHSOGGSKLKEVSLIFRFGPCSKNDKNGEDQIISKVIELLK FESDQIGREIKDKALMSFQDGLSYSTFKITFDLFRVDYRGESEVNRKRGKAAALA IDAIKALTIDNHPNRMRLGFGTKYLYREFYFSPVOONGMERILGISHEVD"
BASE COUNT	427 a 393 c 420 g 338 t
ORIGIN	
Query Match	38.8%; Score 294.6; DB 11; Length 1578;
Best Local Similarity	69.6%; Pred. No. 1.9e-74;
Matches 482; Conservative 0; Mismatches 184; Indels 27; Gaps 5;	
1 atgtgtagaccacagtggtgtgacctctgtaagaatacccccctagatatgtgacctaac 60	
216 ATGTGAGCGACCAATGTGTATGACTCGAAGACATTCCTCGAGAGATGATCCAAAC 275	
61 accatagaattcaaatccctcgcctacacacacagacatcatgtcttcaagagaccct 120	
276 AGCATAGATTCAAAATCTCGCCTTCTTACGCCAGACACCATATGCTTTCAGAACACCCCG 335	
121 gctctcttcaccacaaagctgtgagacaagaagtgtgtcccaagagggccttggggaat 180	
336 GCTGTCTTCTCGGCCCAAGCTCTCCAGAACMAAGAGTCTGTCCCAAGAGCCCTGGGACT 395	
181 tgttcagcaaatgactatgagacagagtgltcatatgaccttgcagaataatcccaatcagt 240	
396 TGTGTTTCACTGATTCCTGACACAGATATCATTTGCTTGGCAGAGATTCCCTCCAGC 452	
241 gagaagagccataacctcttgcaagaanaagttcttcttgaaagcatctcttgtagtagtg 300	
453 GAAAGAGACATCAAGCTTGGGCAAGAGAAGTCTTCTTGAGAACACCTTCCAGAGTGGCC 512	
301 gaga---aggaagattcgcagagacagcgcctgcacaagtgctctgcagaggtcaaaagcag 357	
513 GAAAGAGAGAGAGCGCTCGCGAGCTCCCAAGAGAGATCGAGACTCAGGGTCTCTCAAGGCC 572	
358 ttggaataccaagattcg-----caagcagcagtggttcagagtgtcttctc 405	
573 CCTTCCCGGTAGAGCGGCGACAGTGGCTTCACAAACACACACTGAGGCCCTCTCAGC 632	
406 aaagttagagcagtgcttggaacatgaagctgtgagccccaagtaattcatatgccaac 465	
633 AGTGTGAGAGAGCCCTCGAGAGATGAATTTGGATTTCCAAAATGGCTTATTATGGCCAAC 692	
466 cagtagctaaatattgtattatccctgycgaacacacacaagcgcagcaggaagagcttc 525	
693 AGAGTGGCTGAATATTGTTTACTCTCGGCACACCAACAGATGTCTATCCACACACGAGGAGA 752	

QY 526 aagtcacaaagagatttctgaactgaggtctctctccacgctcccaagccagctgct 585
 Db 753 ACCAACCTCAAGAGAGAGGCTCTCGAGATCTGTGACTTCAGATTGGAAGGAC-----CMT 807
 QY 586 gtagcttcaagtctcaagaagatgaaagaacaatactagccaatatttgagctg 645
 Db 808 GTCACCTCTAAG-----ATAAAGATGAGAGAACCAAAATAAATAGCAAGATGTTGAGCTG 863
 QY 646 ctgaatatctcagagagatcagcttgagaaagaag 678
 Db 864 CTGAATTTCTCGGGGATCTGAGTGGAGAGAG 896
 RESULT 12
 LOCUS AK016670
 DEFINITION AK016670 2222 bp mRNA linear HTC 19-JAN-2002
 ACCESSION AK016670
 VERSION AK016670.1 GI:12855533
 RECORDS 1
 FEATURES
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (sites)
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE 2 (sites)
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 REFERENCE 3 (sites)
 AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Konno, H., Akiyama, J., Nishi, K., Kusunagi, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Yoneda, Y., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, K.,
 Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multiplexed sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861
 REFERENCE 4 (sites)
 AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
 FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5 (bases 1 to 2222)
 AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,
 Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,
 Carinini, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T.,
 Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F.,
 Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawa, M., Koya, S.,
 Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koyama, S.,
 Kurahara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,
 Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,
 Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
 Schmitt, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,
 Sobabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamashita, I.,

Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and
 Hayashizaki, Y.
 Direct Submission
 Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration and Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
 url: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)
 Please visit our web site (http://genome-gsc.riken.go.jp/) for
 further details.
 CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in RIKEN
 Genomic Sciences Center and Genome Science Laboratory in RIKEN
 Division of Experimental Animal Research in RIKEN contributed to
 prepare mouse tissues. First strand cDNA was primed with a primer
 [5' GAGAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCC 3']. cDNA was
 prepared by using trehalose thermo-activated reverse transcriptase
 and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot - 10.0 and subtraction to
 Rot - 100.0. Second strand cDNA was prepared with the primer
 adapter of sequence [5'
 GAGAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCC 3']. cDNA was cleaved
 with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb
 was selected before cloning. Vector: a modified pBluescript KS(+) after
 bulk excision from lambda phage. Cloning sites, 5' end: SalI;
 3' end: BamHI. Host: DH10B.
 Location/Qualifiers
 1..2222
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="MGD:MG1:1897093"
 /db_xref="taxon:10090"
 /clone="4933405K19"
 /sex="male"
 /tissue-type="testis"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 250..1236
 /note="BCL2-like apoptosis inhibitor containing protein
 data source: InterPro, source key: IPR002475, evidence: ISS
 putative"
 /codon_start=1
 /protein_id="BAB30370.1"
 /db_xref="GI:12855534"
 /translation="MCSRSTVYDLEIPLEDPPNTEKILAFYARHAFKKNPAYS
 PKLSRTSLSKALGTWSDTSDTQVSLPCRGSPSSSEKNTSLCKKSSWTLFRVAKE
 EGLSPSKREINAGQPGFPEYRSGFNHNPRLSVEORLESSEVDSKVAICNLR
 VAELVSWPDPDVLHSGGSLKLEKERSYLFYRFGPCDSKRDGDDIISKVELLK
 FSGDOLREIKRDKALMSFDGSLSTFEKTTDLFLRDVDRGSEVAVARQKAYALA
 IDAIKLTATIDNHNMNRLRGCTKRLREYFVWQNGWEKILGSHREVD"
 2205..2210
 /note="putative"
 2222
 /note="putative"
 polyA_signal
 2222
 polyA_site
 2222
 /note="putative"
 BASE COUNT 601 a 565 c 551 g 505 t
 ORIGIN
 QY Query Match 38.8%; Score 294.6; DB 11; Length 2222;
 Best local Similarity 69.6%; Pred. No. 2.2e-74;
 Matches 482; Conservative 0; Mismatches 184; Indels 27; Gaps 5;
 QY 1 atgtgtgacacagtggtgtgtgacctggaagaatccccctagatgatgaactaac 60
 Db 250 ATGTCAGACACACAGTGTATGATCCCTGGAGACATTCCTCGAGATGATCCAAAC 309
 QY 61 accatgaattcaaatctcctgacctaccacacacacatcatgcttcaagacaccc 120
 Db 310 AGCATGAGTTCAAAATCTGCGCTTCTTACGCCAGACACATGCTTCAAGAACCCCG 369
 QY 121 gctctcttcacaaagctgctgagaacaagaagtttcccaagaggggcctg99aat 180

```

Db 370 GCTGCTCTCTCCCAAGCTCTCCAGAACAGAGAGCTGCTCCCAAGAGCCCTGGGACT 429
Oy 181 tcttcagaagaatgagtcagacagagtgctatgctctgcagaatctccatcaagt 240
Db 430 TGGT---CACTGATTCCTGGACACAGATTCATTCCTGACAGAGGTTCCTCCCTCAGC 486
Oy 241 gagaagccataaactctgcaagaagaagctctcttgagaagcatctcttgagtagtg 300
Db 487 GAAAGAACATCAGCTTGGGCAAGAGAGTCTTCTTGAGAACACTCTTCAGGCTGCC 546
Oy 301 gaga---aggaagatcgcagacgcctgcagagctctgcagagtgctcagagtgcaagagc 357
Db 547 GAGAGGAGGAGAGGCTGCTCCAGCTCCCAAGAGAGATCGAGCTCAGGGTCTCAGGGC 606
Oy 358 ttggaataccaaagtctg-----caagcagcagagtgctcagagtgctcttct 405
Db 607 CCCCCTCCCGGTAGACCGGACAGAGTGCTTCCAAACACGACCTGGCCAGGTCTCTGAGC 666
Oy 406 aacgltgagcagtgcttgagacatgaagctgtgagcccaagtcattctccatctgcaac 465
Db 667 AGTGTGGAGACAGCGCTGAGAGTGAAGTGTGATTCCTCAAAAGTGTGATTCCTCAAC 726
Oy 466 cgaatagctgaatgttctatctctgcccacacacacacacacacacacacacacacac 525
Db 727 AGAGTGTGCTGAATGTCTTACTCTGCGCACACACAGATGTCTCAGACGACGAGGAGA 786
Oy 526 aagtcacaaagatttctgaactgagtgctctctctccatctcagatccaaagcagtgct 585
Db 787 AGCAAGCTCAAGAGAGAGGCTCTCGAGATCTGTACTTCAAGTTTGAAAGAC---CTT 841
Oy 586 gtagcttcaagttctaaagaatgaagaagaacaatactaacaataatgttgagctg 645
Db 842 GTGACTCTAAAG---AATTAAGATGAGAGAACCAATATAGCAAGATTTGTTGAGCTG 897
Oy 646 ctgaatatctcagagatcagtgatggaagaag 678
Db 898 CTGAATTTCTCGGGGATCAGTTGGAGAGAG 930

```

```

RESULT 13
BI463753 776 bp mRNA linear EST 21-AUG-2001
LOCUS 603203445F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5269331 5',
DEFINITION mRNA sequence.
ACCESSION BI463753
VERSION BI463753.1 GI:15254409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 776)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaapb-r@mail.nih.gov
Tissue Procurement: Miklos Palcovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1679 row: 1 column: 12
High quality sequence stop: 685.
Location/Qualifiers
1. 776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5269331"

```

FEATURES

source

```

/clone.lib="NIH_MGC_97"
/lab_host="DHL0B"
/Note="Organ: testis; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcagag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to Rof 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 243 a 180 c 191 g 162 t
ORIGIN

```

```

Query Match 36.4% Score 276.2; DB 10; Length 776;
Best Local Similarity 91.2% Pred. No. 3.3e-69;
Matches 340; Conservative 0; Mismatches 23; Indels 10; Gaps 4;
Oy 1 atgtgtgagccagtggtgtgacatggaagaatcccttagatgatgaactaac 60
Db 400 ATGTGTAGCACACAGTGGGTGTGACCTGGAGAAATCCCTAGATGATGATCACTAAAC 459
Oy 61 accatgaattcaaatccctgcctactacacacagacatcatgtcttaagaagacccct 120
Db 460 ACCATGAAATTCAAATCCTCGCTACTACACACAGATCATGTCTTCAAGACACCCCT 519
Oy 121 gctctcttcacaaagctgtctgagacaaagattgtctccagagggcctgaggaaat 180
Db 520 GCTCTCTTCTCACCAAGCTGTGAGAAAGAGATTGTGCCAGAGGGGCTGGGGAAAT 579
Oy 181 tcttgagcaatgaatgacatgagacagagtgatc-ggctctgcagaatctccatccag 239
Db 580 TGTTCAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639
Oy 240 tgaaga-ggcataaacctctgcaagaaga---agctctcttgagaagcatctcttgag 295
Db 640 TGAGAACGGCCATTAACCTTGGCAACGACAGAGTCTTCTTGGGAAGCCATCTTGG 699
Oy 296 tagt-----ggagaaggaagatcgcagagcagcctgcagagtgctctgcaagtgca 350
Db 700 CAGTAGTTGGAGAACCGAATTCGACAGACGCGCTGCCAAAGCTCTGTGTCAGAGTCA 759
Oy 351 aaggaagcttgaa 363
Db 760 AAGGACCGTGGGA 772

```

```

RESULT 14
BI453888 920 bp mRNA linear EST 21-AUG-2001
LOCUS 603174765F1 NCI_CGAP_Man5 Mus musculus cDNA clone IMAGE:5253808 5',
DEFINITION mRNA sequence.
ACCESSION BI453888
VERSION BI453888.1 GI:15244544
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 920)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaapb-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

Plate: L1641 row: b column: 17
High quality sequence stop: 738.

FEATURES

Location/Qualifiers

1. 920

/organism="Mus musculus"

/strain="C57/B6"

/db_xref="taxon:10090"

/clone="IMAGE:5253808"

/clone_1ib="NCI CGAP Mam5"

/tissue_type="tumor, gross tissue"

/dev_stage="7 months"

/lab_host="DH10B"

/note="Organ: mammary; Vector: pCMV-SPOK6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by life technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"

BASE COUNT 244 a 254 c 252 g 170 t

ORIGIN

Query Match 32.5%; Score 246.6; DB 10; Length 920;
Best Local Similarity 71.8%; Pred. No. 1.4e-60;
Matches 371; Conservative 0; Mismatches 129; Indels 17; Gaps 3;

1 atgttagaccagtggtgtgacccctggaagaatccccctagatgatgacctaac 60

158 ATGTGACGACGACCGTGTGTGACCTGGAAGACATTCCTCCGAGGATGATGCCAAG 217

61 accatagaattcaaaatcctgcctactacacacacacacacacacacacacac 120

218 AACATGAGTTCATAAATCCGCTTCACGACACACATGCTTCAGAACACCCCG 277

121 gctctctctaccacaaagcgtcgtgagaagaagttgtccacagaaggcctggggaat 180

278 GCTGCTTCTGCTCCCAAGCTCTCCAGAACAGAGTGTGTGCGAAGAACTGTGGGACT 337

181 ttttcagcaaatgagtcagacagaggtgtcagtcctgacagaattcccaatcagt 240

338 TGGT---CAACTGATTCCTGAGACACAGTATGCTGCTGACAGAGTTCCTCCACAG 394

241 gagaagcacaataacctgtgacagaagaagttctcttggaagcattcttgagtagtg 300

395 GAAAGAACATCATGCTGTGGGCAAGAAAGTCTCTTGGAGAACACTCTTCAGGTGGCC 454

301 gagaagaagaatt---cgagagcagcctgccaaggtctctgctcaggtcagaagagc 357

455 GAGAGGAGGAGGAGCCGACCGAGCTCCCAAGAGATCCAGACTCAGAGGTCTCAGGGC 514

358 ttggaataccaagatttcagacag-----ccagcagtggtccaggtgtcttcta 406

515 CCTTCCCGGTAGAGCGGAGAGTGCTTCCACAAACAGACACTGCGCCAGGTCTGAGCA 574

407 agctggaagcagtgctcttggaacatgaagctgtgagcccaaaagtcatttccatgccaaac 466

575 GGTGTGAGCAGCGCCCTGTGAGAGTGAAGTGTGATTCCTCAAAAGTGGCTGTATTGCCAACA 634

467 gactagcgaattgtttattcctgcgcacacacacaca 503

635 GAGTGGCTGAATCGTTACTCTCTGCGCCACGACAGA 671

RESULT 15

LOCUS AA536718 652 bp mRNA linear EST 29-JUL-1997

DEFINITION vj88a02.r1 Knowles Solter mouse 2 cell Mus musculus cDNA clone

IMAGE:944138 3, mRNA sequence.

ACCESSION AA536718

VERSION AA536718.1 GI:2282711

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 652)

AUTHORS

Maria M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

Contact: Maria M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MGI:540994

High quality sequence stop: 481.

FEATURES

Location/Qualifiers

1. 652

/organism="Mus musculus"

/strain="B6D F1/7"

/db_xref="taxon:10090"

/clone="IMAGE:944138"

/clone_1ib="Knowles Solter mouse 2 cell"

/tissue_type="embryo"

/dev_stage="2-cell"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pBluescribe (modified); Site:1: MluI; Site:2: SalI; Cloned unidirectionally from

RNA prepared from 13,500 2-cell stage embryos. Primer:

SalI(dT): 5'-GGTGGACGCTGACCGCTTTTCTTTTCTTTT-3'. CDNA

were cloned into the MluI/SalI sites of a modified

pBluescribe vector using commercial linkers (NEB).

Average insert size: 1.2 kb.

BASE COUNT 170 a 180 c 178 g 123 t 1 others

ORIGIN

Query Match 32.5%; Score 246.4; DB 9; Length 652;
Best Local Similarity 72.3%; Pred. No. 1.4e-60;
Matches 381; Conservative 0; Mismatches 127; Indels 19; Gaps 4;

1 atgttagaccagtggtgtgacccctggaagaatccccctagatgatgacctaac 60

87 ATGTGACGACGACCGTGTGTGACCTGGAAGACATTCCTCCGAGGATGATGCCAAG 146

61 accatagaattcaaaatcctgcctactacacacacacacacacacacacacac 120

147 AGCATAGAGTTCATAAATCCGCTTCACGACACACCATGCTTCAAGAACACCCCG 206

121 gctctctctaccacaaagcgtcgtgagaagaagttgtccacagaaggcctggggaat 180

207 GCTGCTTCTGCTCCCAAGCTCTCCAGAACAGAGTGTGTGCCAGAACGCCCTGGGACT 266

181 ttttcagcaaatgagtcagacagaggtgtcagtcctgacagaattcccaatcagt 240

267 TGGT---CAACTGATTCCTGAGACACAGTATGCTGCTGACAGAGTTCCTCCAGC 323

467 gactagcgaattgtttattcctgcgcacacacacaca 503

635 GAGTGGCTGAATCGTTACTCTCTGCGCCACGACAGA 671

RESULT 15

LOCUS AA536718 652 bp mRNA linear EST 29-JUL-1997

DEFINITION vj88a02.r1 Knowles Solter mouse 2 cell Mus musculus cDNA clone

IMAGE:944138 3, mRNA sequence.

ACCESSION AA536718

VERSION AA536718.1 GI:2282711

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 05:05:29 ; Search time 302.61 Seconds
(without alignments)
4306.328 Million cell updates/sec

Title: US-09-771-961-3
Perfect score: 759
Sequence: 1 atgtgttagaccagttgtgtg.....gtttgatgcctgatttga 759

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	759	100.0	759	22	AA013236
2	759	100.0	2132	22	AA013237
3	757.4	99.8	954	22	AAH22583
4	678	89.3	984	22	AA013235
5	676.4	89.1	1179	22	AAH22582
6	485.6	64.0	1917	23	AA050453
7	429	56.5	632	22	AA01348
8	388	51.1	388	22	ABA47100
9	388	51.1	388	22	ABA64982

10	388	51.1	388	22	ABA32090	Probe #10556 for g
11	388	51.1	388	22	AAK13408	Human brain expres
12	388	51.1	388	22	AAK39145	Human bone marrow
13	388	51.1	388	22	AAI19954	Probe #9887 for ge
14	388	51.1	388	22	AAI45152	Probe #13838 used
15	388	51.1	388	22	AAI05666	Probe #5657 used t
16	388	51.1	388	22	ABA52358	Human foetal liver
17	317	41.8	466	22	ABA22153	Probe #619 for gen
18	317	41.8	466	22	AAK00629	Human brain expres
19	317	41.8	466	22	AAK26079	Human bone marrow
20	317	41.8	466	22	AAI10708	Probe #641 for gen
21	317	41.8	466	22	AAI31966	Probe #652 used to
22	317	41.8	466	22	AAI00638	Probe #629 used to
23	274.4	36.2	987	22	AAK22670	Mouse Bcl-G polype
24	83	10.9	8922	22	AAK87069	Human Bcl-G polype
25	39.8	5.2	43	22	AAH22593	Human Bcl-G polype
26	39.8	5.2	43	22	AAH22594	Bcl-6s mutagenic p
27	37	4.9	7055	20	AAH26303	Sequence of phage
28	37	4.9	7783	20	AAH26302	Sequence of phage
29	36.2	4.8	954	22	AAH13236	Human Bcl-X-like p
30	36.2	4.8	984	22	AAH22583	Human Bcl-G polyp
31	36.2	4.8	984	22	AAH13235	Human Bcl-X-like p
32	36.2	4.8	1179	22	AAH22582	Human Bcl-G polyp
33	36.2	4.8	2132	22	AAH13237	Human Bcl-X-like p
34	35.8	4.7	10377	24	ABU34220	Human immune syste
35	35.2	4.6	4019	21	AAZ43878	C. vicina arylphor
36	34.6	4.5	6971	20	AAZ26304	Sequence of phage
37	34.2	4.5	600	20	AAV86185	EST clone J635. H
38	34.2	4.5	1001	21	AAH51276	Human GSHS related
39	34	4.5	917	22	AAI94362	Human neuroblastom
40	34	4.5	32195	22	AAH36517	Human cardiovascular
41	34	4.5	32219	22	AAH36516	Human cardiovascular
42	33.8	4.5	2380	18	AAH87981	PCAO-1 modular ve
43	33.8	4.5	2755	18	AAH87967	PCAO-1 modular vect
44	33.8	4.5	3225	22	AAH72847	ECOR1-HindIII lin
45	33.8	4.5	3488	18	AAH87982	PCAO-2 modular ve

ALIGNMENTS

RESULT 1	
AA013236	standard; CDNA; 759 BP.
ID	AA013236
XX	AA013236;
AC	23-OCT-2001 (first entry)
XX	
DT	Human BCL-X-like protein encoding cDNA #2.
DE	
XX	
XX	Human; BCL-X-like protein; therapy; physiological disorder; ss.
KW	
OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
FT	1..759
FT	/*tag= a
FT	/product= "Human BCL-X-like protein #2"
XX	
PN	W0200157213-A2.
PD	
XX	09-AUG-2001.
XX	
PF	31-JAN-2001; 2001WO-US03446.
XX	
PR	04-FEB-2000; 2000US-0180412.
XX	
PA	(LEXI-) LEXICON GENETICS INC.
XX	
PI	Donoho G, Hilbun E, Turner CA, Friedrich G, Abulin A, Zambrowicz B;
PI	Sands AT;
XX	

QY 241 gagaagccataaaccttgcgaagaaagctctcttggaaagcattcttggagtagtg 300
|||||
Db 432 gagaagccataaaccttgcgaagaaagctctcttggaaagcattcttggagtagtg 491
301 gagaagaaagatctgcgaagacagcctgcgaagctctcttgcgaaggttaaggagcttg 360
|||||
Db 492 gagaagaaagatctgcgaagacagcctgcgaagctctcttgcgaaggttaaggagcttg 551
361 gaataccaagatctgcgaagacagcctgcgaagctctcttgcgaaggttaaggagcttg 420
|||||
Db 552 gaataccaagatctgcgaagacagcctgcgaagctctcttgcgaaggttaaggagcttg 611
421 ttggagcatgaagctgtgcgaagcccaagcttccatctgcgaagccgaagtagctgaatt 480
|||||
Db 612 ttggagcatgaagctgtgcgaagcccaagcttccatctgcgaagccgaagtagctgaatt 671
481 gttattctctgcgaagcccaagcagccagcagcagcagcagcagcagcagcagcagcagc 540
672 gttattctctgcgaagcccaagcagccagcagcagcagcagcagcagcagcagcagcagc 731
QY 541 ttgttaactgaggtgtctctctctctccatccagcagcagcagcagcagcagcagcagcagc 600
|||||
Db 732 ttgttaactgaggtgtctctctctctccatccagcagcagcagcagcagcagcagcagcagc 791
QY 601 aagaagatgaagaagaacaatactagccaagaattgttgcagctgtgcgaataatcagga 660
|||||
Db 792 aagaagatgaagaagaacaatactagccaagaattgttgcagctgtgcgaataatcagga 851
QY 661 gatcaattggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 720
|||||
Db 852 gatcaattggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 911
QY 721 caagggtttccacagagtggttgatgagcctgcatlta 759
|||||
Db 912 caagggtttccacagagtggttgatgagcctgcatlta 950

RESULT 3
AAH22583
ID AAH22583 standard; cDNA; 954 BP.
AC AAH22583;
XX
DT 07-SEP-2001 (first entry)
XX
DE Human Bcl-GS polypeptide encoding cDNA.
KW Bcl-G; cancer; cancer therapy; oncogene; apoptosis; Bcl-GS; cytosolic;
antiapoptotic; chromosome 12p12.3; human; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 196..954
FT /tag= a
FT /product= "Bcl-GS"
XX
XX WO200144282-A2.
XX PD 21-JUN-2001.
XX PF 13-DEC-2000; 2000MO-US33793.
XX PR 14-DEC-1999; 99US-0461641.
XX PA (BURN-) BURHAM INST.
XX PI Reed JC, Godzik A;
XX DR WPI: 2001-398125/42.
XX P-PSDB: AAB85167.
XX

PT Novel polynucleotide encoding Bcl-G polypeptide, useful for modulating
PT apoptosis, and for diagnosing and treating cancer
XX
PS Claim 3; Fig 3; 117pp; English.
XX
CC The invention relates to Bcl-G polypeptides and nucleic acids encoding
CC them. The Bcl-G polypeptides can be expressed by standard recombinant
CC methodology. Bcl-G oligonucleotides (or its anti-sense strand) and Bcl-G
CC specific antibodies are useful for diagnosing cancer, monitoring cancer
CC therapy or assessing prognosis of patients with cancer. The Bcl-G
CC polypeptides are useful for modulating the activity of an oncogenic
CC polypeptide. They are useful for identifying modulators, for modulating
CC a level of apoptosis mediated by the Bcl-G polypeptide. A therapeutic
CC composition comprising the Bcl-G polypeptide, polynucleotide or antibody
CC is useful for treating a pathology characterized by abnormal cell
CC proliferation especially cancer. The present sequence represents a cDNA
CC encoding a Bcl-GS polypeptide.
XX
SQ Sequence 954 BP; 267 A; 237 C; 240 G; 210 T; 0 other;

Query Match 99.8%; Score 757.4; DB 22; Length 954;
Best Local Similarity 99.9%; Pred. No. 9,8e-228;
Matches 758; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atgttaacaccagctggtgtgacacctggaagaatcccccataatgataatgaac 60
Db 196 atgttaacaccagctggtgtgacacctggaagaatcccccataatgataatgaac 255
QY 61 accataagaattcaaaatccctgcctactacacacagatcatgtcttcaagaaccct 120
Db 256 accataagaattcaaaatccctgcctactacacacagatcatgtcttcaagaaccct 315
QY 121 gctcctctccacaaagtgtgtgaaagaagaattgttcccaagaaggccttggaat 180
Db 316 gctcctctccacaaagtgtgtgaaagaagaattgttcccaagaaggccttggaat 375
QY 181 ttttaagaatgaatgcatgacagaggtgtcatgagccttgcaagaattcccaatccagt 240
Db 376 ttttaagaatgaatgcatgacagaggtgtgtcatgagccttgcaagaattcccaatccagt 435
QY 241 gagaagccataaaccttgcgaagaaagctcttgcgaagcattcttgcgaagtagtg 300
Db 436 gagaagccataaaccttgcgaagaaagctcttgcgaagcattcttgcgaagtagtg 495
QY 301 gagaagaaagatctgcgaagacagcctgcgaagctcttgcgaaggttaaggagcttg 360
Db 496 gagaagaaagatctgcgaagacagcctgcgaagctcttgcgaaggttaaggagcttg 555
QY 361 gaataccaagatctgcgaagcagcagtggttgcaggtgtcttcaagctggaagcagtg 420
Db 556 gaataccaagatctgcgaagcagcagtggttgcaggtgtcttcaagctggaagcagtg 615
QY 421 ttggagcatgaagctgtgcgaagcccaagatcattccatctgcgaagcagtagctgaatt 480
Db 616 ttggagcatgaagctgtgcgaagcccaagatcattccatctgcgaagcagtagctgaatt 675
QY 481 gttattctctgcgaagcccaagcagccagcagcagcagcagcagcagcagcagcagcagc 540
Db 676 gttattctctgcgaagcccaagcagccagcagcagcagcagcagcagcagcagcagcagc 735
QY 541 ttgttaactgaggtgtctctctccatccagcagcagcagcagcagcagcagcagcagcagc 600
Db 736 ttgttaactgaggtgtctctctccatccagcagcagcagcagcagcagcagcagcagcagc 795
QY 601 aagaagatgaagaagaacaatactagccaagaattgttgcagctgtgcgaataatcagga 660
Db 796 aagaagatgaagaagaacaatactagccaagaattgttgcagctgtgcgaataatcagga 855
QY 661 gatcaattggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 720
Db 856 gatcaattggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 915

QY 721 cagggtttccacagagatgtttgatgagcctgcttga 759
DB 916 cagggtttccacagagatgtttgatgagcctgcttga 954

RESULT 4
ID AAD13235 standard; cDNA; 984 BP.
AC AAD13235;
XX 23-OCT-2001 (first entry)
DE Human BCL-X-like protein encoding cDNA #1.
XX Human; BCL-X-like protein; therapy; physiological disorder; ss.
KW Homo sapiens.
OS

Key Location/Qualifiers
FT CDS 1..984
F1 /product= "Human BCL-X-like protein #1"

XX MO200157213-A2.
XX 09-AUG-2001.
XX 31-JAN-2001; 2001WO-US03446.
XX 04-FEB-2000; 2000US-0180412.
XX (LEXI-) LEXICON GENETICS INC.
XX Donoho G, Hilbun E, Turner CA, Friedrich G, Abuin A, Zambrowicz B;
PI Sands AT;
XX WPI: 2001-488882/53.
XX P-PSDB; AAE07040.
XX Novel isolated human BCL-X-like polynucleotide, useful in therapeutic,
PT diagnostic and pharmacogenic applications -
PS Claim 2a; Page 30; 33pp; English.

XX The present sequence is a cDNA encoding human BCL-X-like protein.
XX The BCL-X-like polynucleotides are useful in therapeutic, diagnostic
XX and pharmacogenic applications. They are useful for screening drugs
XX effective in the treatment of symptomatic or phenotypic manifestations
XX pertaining to the normal function of protein in the body and also for
XX treating physiological disorders and diseases. The BCL-X-like
XX polynucleotides are useful in conjunction with polymerase chain
XX reaction to screen libraries, isolate clones, to prepare cloning
XX and sequencing templates and as hybridisation probes for assessing
XX gene expression patterns.

XX Sequence 984 BP; 289 A; 230 C; 250 G; 215 T; 0 other;

Query Match 89.3%; Score 678; DB 22; Length 984;
Best Local Similarity 100.0%; Pred. No. 9,4e-203;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgtgttagcaccagtggtgtgacctggaagaatccccctagatgacgaactaac 60
DB 1 atgtgttagcaccagtggtgtgacctggaagaatccccctagatgacgaactaac 60
QY 61 accatagaattcaaatcctgcctactacacacgaacatcgtctcaagaagccct 120
DB 61 accatagaattcaaatcctgcctactacacacgaacatcgtctcaagaagccct 120
QY 121 gctctcttcacacaagcgtcgtgagacagaagttgtccagagggcctgggaat 180
DB 121 gctctcttcacacaagcgtcgtgagacagaagttgtccagagggcctgggaat 180

DB 121 gctctcttcacacaagcgtcgtgagacagaagttgtccagagggcctgggaat 180
QY 181 tgtcagaacaatgagtcacagacagaggtgtcattgagccttgacagaattcccaatcag 240
DB 181 tgtcagaacaatgagtcacagacagaggtgtcattgagccttgacagaattcccaatcag 240
QY 241 gagaagccataaaccttggcagaagaagaagttcttggaaagcattcttggagtagtg 300
DB 241 gagaagccataaaccttggcagaagaagaagttcttggaaagcattcttggagtagtg 300
QY 301 gagaaggaagaattcgcagaagcgccttgcagaagtccttcagaggttcaagaagcgttg 360
DB 301 gagaaggaagaattcgcagaagcgccttgcagaagtccttcagaggttcaagaagcgttg 360
QY 361 gaataccaagaattcgcagaagcagcagtggtccaggtgtcttctaagcttgagacagtcg 420
DB 361 gaataccaagaattcgcagaagcagcagtggtccaggtgtcttctaagcttgagacagtcg 420
QY 421 ttggaagcatgaagctgtggaaccccaagaatcattcattgccaaccgaatagctgaatt 480
DB 421 ttggaagcatgaagctgtggaaccccaagaatcattcattgccaaccgaatagctgaatt 480
QY 481 gtttattccttgccaccccaacgaagcgaagcagagagagccttcaagtcacaaagagatt 540
DB 481 gtttattccttgccaccccaacgaagcgaagcagagagccttcaagtcacaaagagatt 540
QY 541 ttgttaactgagaggtctctctcttccagctcccaagcagcagtgtaagcttaagttct 600
DB 541 ttgttaactgagaggtctctctcttccagctcccaagcagcagtgtaagcttaagttct 600
QY 601 aagaaagatgaagaagaacaaataactaagcacaatgttgaagctgtgtaataattcaaga 660
DB 601 aagaaagatgaagaagaacaaataactaagcacaatgttgaagctgtgtaataattcaaga 660
QY 661 gatcagttggaagaagaag 678
DB 661 gatcagttggaagaagaag 678

RESULT 5
ID AAH22582 standard; cDNA; 1179 BP.
AC AAH22582;
XX 07-SEP-2001 (first entry)
XX Human Bcl-G1 polypeptide encoding cDNA.
DE Human Bcl-G1 polypeptide encoding cDNA.
XX Bcl-G; cancer; cancer therapy; oncogene; apoptosis; Bcl-G1; cytostatic;
KW antiapoptotic; chromosome 12p12.3; human; ss.
XX Homo sapiens.
XX

Key Location/Qualifiers
FT CDS 196..1179.
FT /*tag= a
F1 /product= "Bcl-G1"

XX MO200144282-A2.
XX 21-JUN-2001.
XX 13-DEC-2000; 2000WO-US33793.
XX 14-DEC-1999; 99US-0461641.
XX (BURN-) BURNHAM INST.
XX Reed JC, Godzik A;
XX WPI: 2001-398125/42.
XX P-PSDB; AAB85166.

XX Novel polynucleotide encoding Bcl-6 polypeptide, useful for modulating
PT apoptosis, and for diagnosing and treating cancer -
XX
XX
PS Claim 3; Fig 1; 117pp; English.

The invention relates to Bcl-2 polypeptides and nucleic acids encoding them. The Bcl-2 polypeptides can be expressed by standard recombinant methodology. Bcl-2 oligonucleotides (or its anti-sense strand) and Bcl-2 specific antibodies are useful for diagnosing cancer, monitoring cancer therapy or assessing prognosis of patients with cancer. The Bcl-2 polypeptides are useful for modulating the activity of an oncogenic polypeptide. They are useful for identifying modulators, for modulating a level of apoptosis mediated by the Bcl-2 polypeptide. A therapeutic composition comprising the Bcl-2 polypeptide, polynucleotide or antibody is useful for treating a pathology characterized by abnormal cell proliferation especially cancer. The present sequence represents a cDNA encoding a Bcl-2 polypeptide.

Sequence 1179 BP; 340 A; 285 C; 305 G; 249 T; 0 other;

Query Match	89.1%	Score	676.4	DB	22	Length	1179
Best Local Similarity	99.9%	Pred	3.3e-202				
Matches	677	Conservative	0	Mismatches	1	Indels	0
						Gaps	0

QY	1	atggtgtgcacaaatgtgtgtgacctggaaagaatcccccagatgtatgtgacctaac	60
Db	196	atgtgtgtgcacaaatgtgtgtgtgacctggaaagaatcccccagatgtatgtgacctaac	255
QY	61	accatagaattcaaaatcctctgcctacttaacacagacatactgtcttcgaagaccct	120
Db	256	accatagaattcaaaatcctctgcctacttaacacagacatactgtcttcgaagaccct	315
QY	121	gctctcttcacaaagaagctgtctgagaaagaagttgttccagaagggtgcctgggaat	180
Db	316	gctctcttcacaaagaagctgtctgagaaagaagttgttccagaagggtgcctgggaat	375
QY	181	tgttcagcaaatgtgattcgtgacagaaagtgttcattgaccttgacagaaatcccaatcagt	240
Db	376	tgttcagcaaatgtgattcgtgacagaaagtgttcattgaccttgacagaaatcccaatcagt	435
QY	241	gagagaagccctaaaccccttgcagaaagaattctcttggaaagatctcttggagtgtg	300
Db	436	gagagaagccctaaaccccttgcagaaagaattctcttggaaagatctcttggagtgtg	495
QY	301	gagaagaagaatttcgcagagaacagcctctgcacaaagtctctctgcagaggtccaagaagcttg	360
Db	496	gagaagaagaatttcgcagagaacagcctctgcacaaagtctctctgcagaggtccaagaagcttg	555
QY	361	gaataccaagaatttcgcagagccagccagttgttccaggtgtctcttccaaagtgtgagcagttgc	420
Db	556	gaataccaagaatttcgcagagccagccagttgttccaggtgtctcttccaaagtgtgagcagttgc	615
QY	421	tttgagacatgaagcttgttggagccccaagtcattcttcattgccaacgaagttagctggaatt	480
Db	616	tttgagacatgaagcttgttggagccccaagtcattcttcattgccaacgaagttagctggaatt	675
QY	481	gtttattctctgcccacacacacaagaagcgaccccaagtcagagagcttcaagttccaaagaagatt	540
Db	676	gtttattctctgcccacacacacaagaagcgaccccaagtcagagagcttcaagttccaaagaagatt	735
QY	541	tttgttaactagagttctctctcttccttcacgtccaagcgcaagttgctctgattgattcaagtctt	600
Db	736	tttgttaactagagttctctctcttccttcacgtccaagcgcaagttgctctgattgattcaagtctt	795
QY	601	aagaaagaatgaagaagaacaataactgtgcacaaattgtttgagctgtcgaataattccagga	660
Db	796	aagaaagaatgaagaagaacaataactgtgcacaaattgtttgagctgtcgaataattccagga	855
QY	661	gattcaagttggaagaagaag 678	
Db	856	gattcaagttggaagaagaag 873	

RESULT 6

AAS90453
ID AAS90453 standard; cDNA; 1917 BP

AC AAS90453 ;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #26257.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss

OS Homo sapiens.

PN W0200175067-A2

PD 11-OCT-2001

PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
DD 03-MAR-2000; 2000US-0540217

XX
DA /HYCCE-) HYCCE TNOXX
DT

XX
XX
WDT: 2001-630363/73

DR P-PSDB; ABG26266.
YY

PT	New isolated poly
PT	diagnostics for

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1; SEQ ID No 26257; 103pp; English

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1917 BP; 553 A; 443 C; 439 G; 482 T; 0 other;

Query Match	64.0%	Score	485.6	DB	23	Length	1917
Best Local Similarity	83.0%	Pred. No.	4.5e-142				
Matches	669	Conservative	0	Mismatches	9	Indels	128
						Gaps	5

Qy 1 atgtgttagcaccacgtgggtg-tgacctggaagaatccccctagatgatgatgacctaa 59
|||||
96 atgtgttagcaccacgtgggtgttgacctggaagaatccccctagatgatgatgacctaa 155

PR	02-OCT-2000	2000US-0236802.
PR	02-OCT-2000	2000US-0237037.
PR	02-OCT-2000	2000US-0237038.
PR	02-OCT-2000	2000US-0237039.
PR	02-OCT-2000	2000US-0237040.
PR	13-OCT-2000	2000US-0239935.
PR	13-OCT-2000	2000US-0239937.
PR	20-OCT-2000	2000US-0240960.
PR	20-OCT-2000	2000US-0241221.
PR	20-OCT-2000	2000US-0241785.
PR	20-OCT-2000	2000US-0241786.
PR	20-OCT-2000	2000US-0241787.
PR	20-OCT-2000	2000US-0241808.
PR	20-OCT-2000	2000US-0241809.
PR	20-OCT-2000	2000US-0241826.
PR	01-NOV-2000	2000US-0244617.
PR	08-NOV-2000	2000US-0246474.
PR	08-NOV-2000	2000US-0246475.
PR	08-NOV-2000	2000US-0246476.
PR	08-NOV-2000	2000US-0246477.
PR	08-NOV-2000	2000US-0246478.
PR	08-NOV-2000	2000US-0246523.
PR	08-NOV-2000	2000US-0246524.
PR	08-NOV-2000	2000US-0246525.
PR	08-NOV-2000	2000US-0246526.
PR	08-NOV-2000	2000US-0246527.
PR	08-NOV-2000	2000US-0246528.
PR	08-NOV-2000	2000US-0246532.
PR	08-NOV-2000	2000US-0246609.
PR	08-NOV-2000	2000US-0246610.
PR	08-NOV-2000	2000US-0246611.
PR	08-NOV-2000	2000US-0246613.
PR	17-NOV-2000	2000US-0249207.
PR	17-NOV-2000	2000US-0249208.
PR	17-NOV-2000	2000US-0249209.
PR	17-NOV-2000	2000US-0249210.
PR	17-NOV-2000	2000US-0249211.
PR	17-NOV-2000	2000US-0249212.
PR	17-NOV-2000	2000US-0249213.
PR	17-NOV-2000	2000US-0249214.
PR	17-NOV-2000	2000US-0249215.
PR	17-NOV-2000	2000US-0249216.
PR	17-NOV-2000	2000US-0249217.
PR	17-NOV-2000	2000US-0249218.
PR	17-NOV-2000	2000US-0249244.
PR	17-NOV-2000	2000US-0249245.
PR	17-NOV-2000	2000US-0249264.
PR	17-NOV-2000	2000US-0249265.
PR	17-NOV-2000	2000US-0249297.
PR	17-NOV-2000	2000US-0249299.
PR	17-NOV-2000	2000US-0249300.
PR	01-DEC-2000	2000US-0250160.
PR	01-DEC-2000	2000US-0250391.
PR	05-DEC-2000	2000US-0251030.
PR	05-DEC-2000	2000US-0251988.
PR	05-DEC-2000	2000US-0256719.
PR	06-DEC-2000	2000US-0251479.
PR	08-DEC-2000	2000US-0251856.
PR	08-DEC-2000	2000US-0251868.
PR	08-DEC-2000	2000US-0251869.
PR	08-DEC-2000	2000US-0251889.
PR	08-DEC-2000	2000US-0251990.
PR	11-DEC-2000	2000US-0254097.
PR	05-JAN-2001	2001US-0259678.
PA		(HUMA-) HUMAN GENOME SCI INC.
XX		Rosen CA, Barash SC, Ruben SM;
XX		WPI: 2001-465570/50.
XX		P-PSDB: AAM95378.
XX		Isolated nucleic acid molecule encoding a reproductive system antigen

PT is used in preventing, treating or ameliorating a medical condition
XX
XX
XX Claim 1; SEQ ID NO 1349; 1297pp + Sequence Listing; English.
CC
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention.
XX
XX Sequence 632 BP; 190 A; 143 C; 166 G; 128 T; 5 other;

Query Match	56.5%	Score 429;	DB 22;	Length 632;
Best Local Similarity	89.4%;	Pred. No. 1.6e-124;		
Matches 505;	Conservative	0;	Mismatches 54;	Indels 6;
				Gaps 4;

Qy 1 atgtgtagcaccacgtygggtgtgcaccttgaagaataccccctagtatgatgtacccaac 60
Db 52 atgtgtagcaccacgtygggtgtgcaccttgaagaataccccctagtatgatgtacccaac 111

Db 112 accatagaattcaaatcctcgctactacacacagacacatgctcttcaagagacacccct 171

Db 172 gctctctctcaccagaagctgtagaacaagaagtttgcaccagaggggctctgggaat 231

232 tgttcagcaaatgtagtcatagcagaggtgtctatgcctgcagaattcccatcagt 291

[illegible]

Db	352	gagaaggaagatctcgacagagacgcctccaaggtctctcgaaggtcaaaagacgttg	411
Oy	361	gaataccaagaattcgcacacagccagcagatggtccaggtgtcttcttaacgtgagcagatgc	420

Dd 412 gaataccaagattcgacacag-cagcagtyg--ccagctgcttctaagtygagcaagtg 4688

QY 421 ttgagcatgaagctgtgagcccaagtcattcattgccaacogafagctgaatt 4800

Db 469 ctctgagcaatgaactgtgagcccaag-catttncattgcaaccagatgactgaaat 527
QY 481 gttattcctggc--caccaccacaagcgaccagcgaggttcaagtccaagaga 538

539 tttttgtaactgaggtctctcctt 563
||||| ||||| || || || ||

RESULT 8

ID	ABA47100 standard; DNA; 388 BP;
XX	
AC	ABA47100;
XX	

DI	01-FEB-2002	(first entry)
XX		
DE		Human breast cell single exon nucleic acid probe #5795.
XX		

am disease, cancer / ss.
XX
OS Homo sapiens.
XX

PN W0200157271-A2.
 XX 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US00662.
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-496933/54.
 DR
 XX
 PS
 CC New spatially-addressable set of single exon nucleic acid probes,
 CC useful for measuring gene expression in sample derived from human
 CC breast, comprises number of single exon nucleic acid probes
 CC
 CC Claim 4; SEQ ID NO 5795; 327pp + sequence listing; English.
 CC
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a single exon nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;
 CC
 CC
 CC Query Match 51.1%; Score 388; DB 22; Length 388;
 CC Best Local Similarity 100.0%; Pred. NO. 1e-111;
 CC Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 46 gatgatgacctaaacacatagatcaaatctctgcctactacacagatcatgtc 105
 CC 1 gatgagacctaaacacatagatcaaatctctgcctactacacagatcatgtc 60
 CC 106 ttcagaagcaccctgtctctctcttcacccaagctgtcgaagaagaagttgtccag 165
 CC 61 ttcagaagcaccctgtctctctcttcacccaagctgtcgaagaagaagttgtccag 120
 CC 166 aggggctctgggaattgttcagaatgatgatcagagaggttcagtccttcag 225
 CC 121 aggggctctgggaattgttcagaatgatgatcagagaggttcagtccttcag 180
 CC 226 aattcccaatccgctgtagaagccataaactctggcaagaagaagttcttggaaaga 285
 CC 181 aattcccaatccgctgtagaagccataaactctggcaagaagaagttcttggaaaga 240
 CC 286 ttccttgagagtaatgagaaagaaatctgcagagcagcctgccaaggtctctcag 345
 CC 241 ttccttgagagtaatgagaaagaaatctgcagagcagcctgccaaggtctctcag 300
 CC 346 ggtcaaaagagcgttggaataccaagattcgcacagccagcagtggtccaggtcttct 405

DB 301 ggtcaaaagagcgttggaataccaagattcgcacagccagcagtggtccaggtcttct 360
 QY 406 aacgttgagcagcgttcttggagcatgaag 433
 DB 361 aacgttgagcagcgttcttggagcatgaag 388
 CC
 CC RESULT 9
 CC ABA64982
 CC ID ABA64982 standard; DNA; 388 BP.
 CC XX
 CC ABA64982;
 CC XX
 CC DT 01-FEB-2002 (first entry)
 CC XX
 CC DE Human foetal liver single exon nucleic acid probe #13287.
 CC XX
 CC KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
 CC XX
 CC OS Homo sapiens.
 CC XX
 CC PN W0200157277-A2.
 CC XX
 CC PD 09-AUG-2001.
 CC XX
 CC PF 30-JAN-2001; 2001WO-US00669.
 CC XX
 CC PR 04-FEB-2000; 2000US-0180312.
 CC PR 26-MAY-2000; 2000US-0207456.
 CC PR 30-JUN-2000; 2000US-0608408.
 CC PR 03-AUG-2000; 2000US-0632366.
 CC PR 21-SEP-2000; 2000US-0234687.
 CC PR 27-SEP-2000; 2000US-0236359.
 CC PR 04-OCT-2000; 2000GB-0024263.
 CC
 CC (MOLE-) MOLECULAR DYNAMICS INC.
 CC PA Penn SG, Hanzel DK, Chen W, Rank DR;
 CC PI WPI; 2001-483447/52.
 CC DR
 CC XX
 CC XX Human genome-derived single exon nucleic acid probes useful for
 CC PT analyzing gene expression in human foetal liver -
 CC PS
 CC PS Claim 4; SEQ ID NO 13287; 639pp + sequence listing; English.
 CC XX
 CC CC The invention relates to a single exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 CC Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;
 CC
 CC
 CC Query Match 51.1%; Score 388; DB 22; Length 388;
 CC Best Local Similarity 100.0%; Pred. NO. 1e-111;
 CC Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC
 CC 46 gatgatgacctaaacacatagatcaaatctctgcctactacacagatcatgtc 105
 CC 1 gatgatgacctaaacacatagatcaaatctctgcctactacacagatcatgtc 60
 CC 106 ttcagaagcaccctgtctctcttcacccaagctgtcgaagaagaagttgtccag 165
 CC 61 ttcagaagcaccctgtctctcttcacccaagctgtcgaagaagaagttgtccag 120
 CC 166 aggggctctgggaattgttcagaatgatgatcagagaggttcagtccttcag 225

Db 121 aggggctggggaattgtctcagcaaatgtagtgcagagagtggtcattgcttcaga 180
 |||
 QY 226 aattcccaatccagtggaagaagccataaactctggcaagaagaatctcttggaaaga 285
 |||
 Db 181 aattcccaatccagtggaagaagccataaactctggcaagaagaagaatctcttggaaaga 240
 |||
 QY 286 ttctttgagtagtggaagaagaatctgcagagcagcgtctcccaaggctctctgcag 345
 |||
 Db 241 ttctttgagtagtggaagaagaatctgcagagcagcgtctcccaaggctctctgcag 300
 |||
 QY 346 ggtcagaagagcgttggaatcccaagattccgacagccagcaatgtgtccaggtcttct 405
 |||
 Db 301 ggtcagaagagcgttggaatcccaagattccgacagccagcaatgtgtccaggtcttct 360
 |||
 QY 406 aacgtgagcagtgcttgagcatgaag 433
 |||
 Db 361 aacgtgagcagtgcttgagcatgaag 388
 |||
 RESULT 10
 ID ABA32090
 AC ABA32090 standard; DNA; 388 BP.
 AC ABA32090;
 DT 23-JAN-2002 (first entry)
 DE Probe #10556 for gene expression analysis in human heart cell sample.
 DE Human; gene expression; heart; microarray; vascular system; probe;
 DE cardiovascular disease; hypertension; cardiac arrhythmia;
 DE congenital heart disease; ss.
 OS Homo sapiens.
 OS
 XX WO200157274-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00666.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-488899/53.
 DR
 PT Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX
 PS Claim 4; SEQ ID NO 10556; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart. The
 CC present sequence is one such probe. The probes may be used for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from the human heart via microarrays. By measuring gene expression, the
 CC probes are useful for predicting, diagnosing, grading, staging,
 CC monitoring and prognosing diseases of the human heart and vascular system
 CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
 CC congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIFO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;
 Query Match 51.1%; Score 388; DB 22; Length 388;
 Best Local Similarity 100.0%; Pred. No. 1e-111;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 46 gatgtgacctaaacacccatagattccaatctctgctctactacaccagatcatgtc 105
 |||
 Db 1 gatgtgacctaaacacccatagattccaatctctgctctactacaccagatcatgtc 60
 |||
 QY 106 ttcaagagacccctgctctctctccaccaaagctgtgtgaacaagaagttgtccag 165
 |||
 Db 61 ttcaagagacccctgctctctctccaccaaagctgtgtgaacaagaagttgtccag 120
 |||
 QY 166 aggggctggggaattgttcaagcaaatgtatgtacatgtgacagagtgatgtgcttcaga 225
 |||
 Db 121 aggggctggggaattgttcaagcaaatgtatgtacatgtgacagagtgatgtgcttcaga 180
 |||
 QY 226 aattcccaatccagtggaagaagccataaactctggcaagaagaagaatcttctggaaaga 285
 |||
 Db 181 aattcccaatccagtggaagaagccataaactctggcaagaagaagaagaatcttctggaaaga 240
 |||
 QY 286 ttctttgagtagtggaagaagaatctgcagagcagcgtctcccaaggctctctgcag 345
 |||
 Db 241 ttctttgagtagtggaagaagaatctgcagagcagcgtctcccaaggctctctgcag 300
 |||
 QY 346 ggtcagaagagcgttggaatcccaagattccgacagccagcaatgtgtccaggtcttct 405
 |||
 Db 301 ggtcagaagagcgttggaatcccaagattccgacagccagcaatgtgtccaggtcttct 360
 |||
 QY 406 aacgtgagcagtgcttgagcatgaag 433
 |||
 Db 361 aacgtgagcagtgcttgagcatgaag 388
 |||
 RESULT 11
 ID AAK13408
 AC AAK13408 standard; DNA; 388 BP.
 AC AAK13408;
 DT 05-NOV-2001 (first entry)
 DE Human brain expressed single exon probe SEQ ID NO: 13399.
 DE Human; brain expressed exon; gene expression analysis; probe;
 DE microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 DE epilepsy; cancer; ss.
 OS Homo sapiens.
 OS
 XX WO200157275-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00667.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 DR WPI; 2001-483446/52.
 DR
 XX

XX XX WO200157278-A2.
 XX XX 09-AUG-2001.
 XX XX 30-JAN-2001; 2001WO-US00670.
 PR XX 04-FEB-2000; 2000US-0180312.
 PR XX 26-MAY-2000; 2000US-0207456.
 PR XX 30-JUN-2000; 2000US-0608408.
 PR XX 03-AUG-2000; 2000US-0632366.
 PR XX 21-SEP-2000; 2000US-0234687.
 PR XX 27-SEP-2000; 2000US-0236359.
 PR XX 04-OCT-2000; 2000GB-0024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PT WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 25; SEQ ID NO 9887; 487bp; English.
 CC The present invention relates to human single exon nucleic acid probes
 CC (SENP). The present sequence is one such probe. The SENPs are derived
 CC from human Hela cells. The SENPs can be used to produce a single exon
 CC microarray, which can be used for measuring human gene expression in a
 CC sample derived from human cervical epithelial cells. By measuring gene
 CC expression, the probes are therefore useful in grading and/or staging
 CC of diseases of the cervix, notably cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;
 SQ

Query Match 51.1%; Score 388; DB 22; Length 388;
 Best Local Similarity 100.0%; Pred. NO. 1e-111;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gatgatgacctaaccacatgaattcaaatccctgcctactacacagacatcatg 105
 D 1 gatgatgacctaaccacatgaattcaaatccctgcctactacacagacatcatg 60
 QY 106 ttcaagagcaccctctctctctcaccacaaagctgctgagaaagaagttgtccag 165
 D 1 ttcaagagcaccctctctctctcaccacaaagctgctgagaaagaagttgtccag 120
 QY 61 ttcaagagcaccctctctctctcaccacaaagctgctgagaaagaagttgtccag 120
 QY 166 aggggctggggaattgttgaagcaatgagtcagacagagtggtcattgaccttcaga 225
 D 121 aggggctggggaattgttgaagcaatgagtcagacagagtggtcattgaccttcaga 180
 QY 226 aattcccaatccagctgagaaagcacaataacttggcaagaagaatcttcttggaaaga 285
 D 181 aattcccaatccagctgagaaagcacaataacttggcaagaagaatcttcttggaaaga 240
 QY 286 ttcttggagtagtgagaaagaagatcgacagacagcctgccaaggtctctgtccag 345
 D 241 ttcttggagtagtgagaaagaagatcgacagacagcctgccaaggtctctgtccag 300
 QY 346 ggtcacaagacgttggagaaagatcgacacagccagcagcagtggtccaggtcttctc 405
 D 301 ggtcacaagacgttggagaaagatcgacacagccagcagcagtggtccaggtcttctc 360
 QY 406 aacgtgagacagtgcttggagacatgaag 433
 D 361 aacgtgagacagtgcttggagacatgaag 388

RESULT 14
 ID AAI45152 standard; DNA; 388 BP.
 AC AAI45152;
 DT 17-OCT-2001 (first entry)
 DE Probe #13838 used to measure gene expression in human placenta sample.
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder; ss.
 OS Homo sapiens.
 EN WO200157272-A2.
 PD 09-AUG-2001.
 PF 30-JAN-2001; 2001WO-US00663.
 PR XX 04-FEB-2000; 2000US-0180312.
 PR XX 26-MAY-2000; 2000US-0207456.
 PR XX 30-JUN-2000; 2000US-0608408.
 PR XX 03-AUG-2000; 2000US-0632366.
 PR XX 21-SEP-2000; 2000US-0234687.
 PR XX 27-SEP-2000; 2000US-0236359.
 PR XX 04-OCT-2000; 2000GB-0024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PT WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 25; SEQ ID NO 13838; 654bp; English.
 CC The present invention relates to single exon nucleic acid probes (SENP).
 CC The present sequence is one such probe. The probes are useful for
 CC producing a microarray for predicting, measuring and displaying gene
 CC expression in samples derived from human placenta. The probes are useful
 CC for antenatal diagnosis of human genetic disorders.
 CC
 XX Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;
 SQ

Query Match 51.1%; Score 388; DB 22; Length 388;
 Best Local Similarity 100.0%; Pred. NO. 1e-111;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gatgatgacctaaccacatgaattcaaatccctgcctactacacagacatcatg 105
 D 1 gatgatgacctaaccacatgaattcaaatccctgcctactacacagacatcatg 60
 QY 106 ttcaagagcaccctctctctcaccacaaagctgctgagaaagaagttgtccag 165
 D 61 ttcaagagcaccctctctctcaccacaaagctgctgagaaagaagttgtccag 120
 QY 61 ttcaagagcaccctctctctcaccacaaagctgctgagaaagaagttgtccag 120
 QY 166 aggggctggggaattgttgaagcaatgagtcagacagagtggtcattgaccttcaga 225
 D 121 aggggctggggaattgttgaagcaatgagtcagacagagtggtcattgaccttcaga 180
 QY 226 aattcccaatccagctgagaaagcacaataacttggcaagaagaatcttcttggaaaga 285
 D 181 aattcccaatccagctgagaaagcacaataacttggcaagaagaatcttcttggaaaga 240
 QY 286 ttcttggagtagtgagaaagaagatcgacagacagcctgccaaggtctctgtccag 345
 D 241 ttcttggagtagtgagaaagaagatcgacagacagcctgccaaggtctctgtccag 300

QY 346 ggtcaaggacgttggaataccagaattcgacagccagctggttcagggtcttct 405
 |||
 Db 301 ggtcaaggacgttggaataccagaattcgacagccagctggttcagggtcttct 360
 QY 406 aacgttgagcagctgcttgagcatgaag 433
 |||
 Db 361 aacgttgagcagctgcttgagcatgaag 388

RESULT 15
 AA105666 standard; DNA; 388 BP.
 ID AA105666
 AC AA105666;
 XX
 DT 09-OCT-2001 (first entry)
 DE Probe #5657 used to measure gene expression in human breast sample.
 XX
 KW Probe; human; breast disease; breast cancer; development disorder; sw;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OM Homo sapiens.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US00661.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-476286/51.
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 in a human breast -
 XX
 PS Claim 25; SEQ ID NO 5657; 322bp; English.
 XX
 SQ The present invention relates to novel single exon nucleic acid probes.
 The present sequence is one such probe. The probes are useful for
 measuring human gene expression in a human breast sample, where the probe
 hybridises at high stringency to a nucleic acid expressed in the human
 breast. The probes are useful for predicting, diagnosing, grading,
 staging, monitoring and prognosing diseases of the human breast,
 particularly those diseases with polygenic aetiology. The diseases
 include: breast cancer, disorders of development, inflammatory diseases
 of the breast, fibrocystic changes, proliferative breast disease and
 non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pft_sequences.
 XX
 SQ Sequence 388 BP; 113 A; 91 C; 99 G; 85 T; 0 other;

Query Match 51.1%; Score 388; DB 22; Length 388;
 Best Local Similarity 100.0%; Pred. No. 1e-111;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 gatgatgacttaaacaccatagaattcaaatcctcgctactataccagacatcatgtc 105
 |||
 Db 1 gatgatgacttaaacaccatagaattcaaatcctcgctactataccagacatcatgtc 60

QY 106 ttcaagacacccctgctctctctctcaccacaagctgtgagaacaagaattgtccag 165
 |||
 Db 61 ttcaagacacccctgctctctctctcaccacaagctgtgagaacaagaattgtccag 120
 QY 166 aggggctggggaattgttcagcaaatgagtcataggacagagtgatcagccttcaga 225
 |||
 Db 121 aggggctggggaattgttcagcaaatgagtcataggacagagtgatcagccttcaga 180
 QY 226 aattcccaatccagtgaggaaggccataaaccttggcaagaaagcttcttggaagca 285
 |||
 Db 181 aattcccaatccagtgaggaaggccataaaccttggcaagaaagcttcttggaagca 240
 QY 286 ttcttgagtagtgaggaagaagatctcgacagcagccttgccaaggtctctgctcag 345
 |||
 Db 241 ttcttgagtagtgaggaagaagatctcgacagcagccttgccaaggtctctgctcag 300
 QY 346 ggtcaaggacgttggaataccagaattcgacagccagcagctggttcagggtcttct 405
 |||
 Db 301 ggtcaaggacgttggaataccagaattcgacagccagcagctggttcagggtcttct 360
 QY 406 aacgttgagcagctgcttgagcatgaag 433
 |||
 Db 361 aacgttgagcagctgcttgagcatgaag 388

Search completed: June 20, 2002, 05:05:37
 Job time: 5981 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 05:05:07 ; Search time 2339.76 Seconds
(without alignments)
6788.406 Million cell updates/sec

Title: US-09-771-961-3

Perfect score: 759
Sequence: 1 atgtgtacaccagtggtggtggtggttgcgtgcgtgattga 759

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb_ba: *
2: gb_hlg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_hlg_hum: *
31: em_hlg_inv: *
32: em_hlg_other: *
33: em_hlgo_inv: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	759	100.0	759	6	AX207694	AX207694 Sequence
2	759	100.0	2132	6	AX207696	AX207696 Sequence
3	757.4	99.8	954	6	AX174785	AX174785 Sequence
4	757.4	99.8	2039	9	AF281255	AF281255 Homo sapi
5	678.8	89.4	1176	9	AY040274	AY040274 Homo sapi
6	678	89.3	984	6	AX207692	AX207692 Sequence
7	676.4	89.1	1179	6	AX174783	AX174783 Sequence
8	676.4	89.1	1886	6	AF281254	AF281254 Homo sapi
9	435.8	57.4	190858	9	AC007537	AC007537 Homo sapi
10	435.8	57.4	194143	2	AC022222	AC022222 Homo sapi
11	435.8	57.4	200499	2	AC091816	AC091816 Homo sapi
12	274.4	36.2	987	6	AX174823	AX174823 Sequence
13	176	23.2	149350	2	AC097248	AC097248 Rattus no
14	175.2	23.1	2877	2	AX026440	AX026440 Homo sapi
15	62.4	8.2	149350	2	AC097248	AC097248 Rattus no
16	60.8	8.0	699	9	HSB329005	HSB329005 Homo sapi
17	41	5.4	125020	9	AF429315	AF429315 Homo sapi
18	39.8	5.2	43	6	AX174816	AX174816 Sequence
19	39.8	5.2	43	6	AX174817	AX174817 Sequence
20	39	5.1	63722	2	AC103740	AC103740 Homo sapi
21	39	5.1	176372	2	AC026030	AC026030 Homo sapi
22	39	5.1	177483	9	AC068233	AC068233 Homo sapi
23	37.8	5.0	197852	10	AL606509	AL606509 Mouse DNA
24	37.4	4.9	54632	2	AC094041	AC094041 Rattus no
25	37.2	4.9	151075	2	AC022594	AC022594 Homo sapi
26	37.2	4.9	166526	2	AC104061	AC104061 Homo sapi
27	37.2	4.9	167132	9	AC069154	AC069154 Homo sapi
28	37	4.9	7055	6	AX006017	AX006017 Sequence
29	37	4.9	7218	6	166494	166494 Sequence 14
30	37	4.9	7783	6	AX006002	AX006002 Sequence
31	36.8	4.8	37895	3	U58748	U58748 Caenorhabdi
32	36.4	4.8	83762	2	AC008707	AC008707 Homo sapi
33	36.4	4.8	92192	2	AC093358	AC093358 Mus muscu
34	36.4	4.8	125020	9	AF429315	AF429315 Homo sapi
35	36.4	4.8	176629	9	AC008641	AC008641 Homo sapi
36	36.4	4.8	205819	2	AC106805	AC106805 Homo sapi
37	36.2	4.8	759	6	AX207694	AX207694 Sequence
38	36.2	4.8	954	6	AX174785	AX174785 Sequence
39	36.2	4.8	984	6	AX207692	AX207692 Sequence
40	36.2	4.8	1176	9	AY040274	AY040274 Homo sapi
41	36.2	4.8	1179	6	AX174783	AX174783 Sequence
42	36.2	4.8	1886	9	AF281254	AF281254 Homo sapi
43	36.2	4.8	2039	6	AF281255	AF281255 Homo sapi
44	36.2	4.8	2132	9	AX207696	AX207696 Sequence
45	36	4.7	25640	2	AC012637	AC012637 Homo sapi

ALIGNMENTS

RESULT	1	759 bp	DNA	Linear	PAT 31-AUG-2001
AX207694	AX207694	Sequence 3 from Patent WO0157213.			
LOCUS	AX207694				
DEFINITION	AX207694				
ACCESSION	AX207694.1	GI:15422375			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
human.					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					

REFERENCE 1 (bases 1 to 759)
Donoho, G., Hilbur, E., Turner, C.A., Friedrich, G.B., Abuln, A.,
Zambrowicz, B., and Sands, A.T.
Human bcl-x-like proteins and polynucleotides encoding the same
Patent: WO 0157213-A 3 09-AUG-2001.
Lexicon Genetics Incorporated (us)
FEATURES
SOURCE
1. 759
/Organism="Homo sapiens"
/db_xref="taxon:9606"
Location/Qualifiers

BASE COUNT	216 a	182 c	185 g	176 t
ORIGIN				

VERSION	AX174785.1	GI:14598301
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 954)	
AUTHORS	Reed,J.C. and Godzik,A.	
TITLE	Bcl-1 polypeptides, encoding nucleic acids and methods of use	
JOURNAL	Patent: WO 0144282-A 3 21-JUN-2001; The Burnham Institute (US)	
FEATURES	Location/Qualifiers	
source	1..954	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	196..954	
CDS	/note="unnamed protein product"	
	/codon_start=1	
	/protein_id="CAC43801.1"	
	/db_xref="GI:14598302"	
	/translation="MCSSTGCDLEIPDDDDLTNTEKLIAYTRHHVKSPPALPS PLIITRSLRSORIGINC SANSMTVEVSPCRNSOSEKINLGKSSKMAKFEVVERE ESQSITPAKYSAGQRTLEIDYDSISQMSKLSINVEQLEHADVRLVISIANVAEL VSWPEQATAGGFKSEKLEIVTEGLSLQDGHVPAVSSSKDDEQILAKIVELLYKY SDSDERKKTAFPIPIPLVDYSIQGFPODGLMACI"	
BASE COUNT	267 a	237 c 240 g 210 t
ORIGIN		
Query Match	99.8%	Score 757.4; DB 6; Length 954;
Best Local Similarity	99.9%	Pred. No. 3.6e-220;
Matches 758; Conservative	0; Mismatches	1; Indels 0; Gaps 0;
1	atgtgtagaccacggttggtgtgtgacctggaagaatccccctagatgatgtgacctaac	60
196	atgtgtagaccacggttggtgtgtgacctggaagaatccccctagatgatgtgacctaac	255
61	accatgaattccaatccctcgctctactacaccagagatcatgtcttccaagagaccct	120
256	accatgaattccaatccctcgctctactacaccagagatcatgtcttccaagagaccct	315
121	gctctcttccaccaaagctgtgtgaaacaagaaglttltgccagagaggcttgggaat	180
316	gctctcttccaccaaagctgtgtgaaacaagaaglttltgccagagaggcttgggaat	375
181	tggttcgcaaatgagtcacatgagacagaggtgtcattgagccttgcagaatccccatc	240
376	tggttcgcaaatgagtcacatgagacagaggtgtcattgagccttgcagaatccccatc	435
241	ggaagagccataaaccttgcagaagaagaatcttcttgaaagatcttcttgtagt	300
436	ggaagagccataaaccttgcagaagaagaatcttcttgaaagatcttcttgtagt	495
301	ggaagaggaatttcgcagagacgcttcgaaggtctctgcagaggttccaagaagctgtg	360
496	ggaagaggaatttcgcagagacgcttcgaaggtctctgcagaggttccaagaagctgtg	555
361	gaataccaagaatttcgcagagacgcttcgaaggtctctgcagaggttccaagaagctgtg	420
556	gaataccaagaatttcgcagagacgcttcgaaggtctctgcagaggttccaagaagctgtg	615
421	ttggagcaatgaagcttgcagagacgcttcgaaggtctctgcagaggttccaagaagctgtg	480
616	ttggagcaatgaagcttgcagagacgcttcgaaggtctctgcagaggttccaagaagctgtg	675
481	gttattctctgcagac	540
676	gttattctctgcagac	735
541	ttgttgaactgaagggtctctctctccagctccaaggccacgtgctgttagcttcaagttc	600
736	ttgttgaactgaagggtctctctctccagctccaaggccacgtgctgttagcttcaagttc	795

QY	601	aagaagaatgaagaagaacaataactagaaccaaatgtttgagctgtctgaataattcaga	660
Db	796	AAGAAAGATGAGAAAGAACCAATACTAGCCAAAATGTTGAGCTGCTGAATAATTCAGCA	855
QY	661	gattcagttggaagaagaaggaacatgccttcattcccatcttggtttgacaccagatc	720
Db	856	GATCAGATTGGAAAAGGACACTGCTTCATCCCATCCCTGTTGGTTGACACAGCATC	915
QY	721	caaggtttccacagatggtttgattgaccttcatttga	759
Db	916	CAGGGTTTTCACAGGATGTTTGATGGCTGCATTTGA	954
RESULT	4		
LOCUS	AF281255	2039 bp	mrna linear PRI 28-JAN-2001
DEFINITION	Homo sapiens apoptosis regulator BCL-G short form (BCLg) mRNA,		
ACCESSION	AF281255	complete cds.	
VERSION	AF281255.1	GI:12584086	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 2039)		
AUTHORS	Guo,B., Godzik,A. and Reed,J.C.		
TITLE	Bcl-9, a novel pro-apoptotic member of the bcl-2 family		
JOURNAL	J. Biol. Chem. 276 (4), 2780-2785 (2001)		
MEDLINE	21264734		
REFERENCE	2 (bases 1 to 2039)		
AUTHORS	Guo,B., Godzik,A. and Reed,J.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-JUN-2000) The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES	Location/Qualifiers		
source	1..2039		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="12"		
gene	1..2039		
	/gene="BCLG"		
	196..954		
	/db_xref="GI:12584087"		
	/translation="MCSSTSGCDLEIPDDDDNLTIEKILAYTFHHFKSPALFS PKILTRLSLRGLGNCSSANESWTEVSPCRSSSEKAIINLGRKSSKAFGEVKE EDSQSPAPKVASQGRRLTLYODSHSOQSLSNVECLHEHVAIPKVISIANVAEI VSWMPPOATQAGFGSKELFVTEGLISFOLGHHVAVSSSKRDEEQIILIVELLY SGDLERKDPAIFIRPLVDTSIQGPQGLMACI"		
CDs	/note="member of Bcl-2 family"		
	/codon_start=1		
	/product="apoptosis regulator BCL-G short form"		
	/protein_id="AAC59794.1"		
	/db_xref="GI:12584087"		
	/translation="MCSSTSGCDLEIPDDDDNLTIEKILAYTFHHFKSPALFS PKILTRLSLRGLGNCSSANESWTEVSPCRSSSEKAIINLGRKSSKAFGEVKE EDSQSPAPKVASQGRRLTLYODSHSOQSLSNVECLHEHVAIPKVISIANVAEI VSWMPPOATQAGFGSKELFVTEGLISFOLGHHVAVSSSKRDEEQIILIVELLY SGDLERKDPAIFIRPLVDTSIQGPQGLMACI"		
BASE COUNT	567 a	475 c	484 g 513 t
ORIGIN			
Query Match	99.8%	Score 757.4	DB 9; Length 2039;
Best Local Similarity	99.9%	Pred. No. 3.9e-220;	
Matches 758; Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	1	atgtgtgacacacagtggtgtgaacctggaagaatacccccctagatgatgatgaac	60
Db	196	ATGTGTAGCACCAATGGGTGTGACCTGGAAGAAATCCCTAGATGATGACCTAAC	255
QY	61	accatagaatcaaatccctgcctcactaacaccagaatcatcatgtcttaagaagacacct	120
Db	256	ACCATGAAATTCAAATTCCTGCGCTACTACACCAACATCATGTCTTCAAGACACCCCT	315
QY	121	gctctctccacacacagctgctgagaacaagaattttcccaagagggcctgggaat	180
Db	316	GCCTCTCTCTCCACCAAGACCTCTGTAGCAACAAATACTTTCTCCAGACGGCCCTCGGCAAT	375

```

QY 181 ttctcagaatgaatgcatgagacagagtgatcagccttgcaagaattcccaatccagt 240
    |||||||
Db 376 TGTTCAGCAATATGATGATGACAGAGGTGTCAAGGCTTGAGAAATTCACATCCAGT 435
    |||||||
QY 241 gagaagccaataaccttggcaagaagaagctctcttggaaagcctctcttggagtagtg 300
    |||||||
Db 436 GAGAAAGCCATTAACCTTGGCAAGAAAGTCTCTTGGAAGCATTTCTTGGAGTAGTG 495
    |||||||
QY 301 gagaagaagaatctgcagagcagccttgcgaagtctctgtcgaaggttcaagaagagcttg 360
    |||||||
Db 496 GAGAAAGCAATTCGCGAGACACGCTGCCAAGGTCTCTGCTCAAGGCTCAAGAGAGCTTG 555
    |||||||
QY 361 gaatacaagattcgcagcagcagcagtggtcagagtgctcttcaacgttggagcagtg 420
    |||||||
Db 556 GAATACCAAGATTTCGACAGCCAGCAGTGTGTCAGGTCTTTTCAAGCTGGAGCAGTGC 615
    |||||||
QY 421 ttggagcatgaagctgttgcaccccaagtcatttccatttgcacacgagtagctgaatt 480
    |||||||
Db 616 TTGAGCATGAGCTGTGAGACCCCAAGTCAATTCATTGGCAACCGAGTAGCTGAATTT 675
    |||||||
QY 481 gttattctctgcacacacacacagcagcagcagcagcagcagcagcagcagcagcagc 540
    |||||||
Db 676 GTTACTCTCGCCACACCAACAGCCAGCAGCAGGAGGCTTCAAGTCCAAAGAGATT 735
    |||||||
QY 541 ttgttaactgaaggtctctctctcagcagcagcagcagcagcagcagcagcagcagcag 600
    |||||||
Db 736 TTTGTACTAGAGGCTCTCTCTTCCAGCTCCAAAGGCCAGTGCCTGTACTTCAAGTTCT 795
    |||||||
QY 601 aagaagaatgaagaagaacaatactagcacaagaattgttgcagctgtgaataatcagga 660
    |||||||
Db 796 AAGAAAGATGAAGAGAAACAATATGCTAGCCAAATTTGTTGAGCTGCTGAATATTTCAGA 855
    |||||||
QY 661 gatcagttgaaagaagaagacacgtgcttcatcccatcttcccttggttgacacccagcag 720
    |||||||
Db 856 GATCAGTTGGAAGAAAGAGACGCTCTTCATCCCATTTCCCTTGTTGAGACACACGACATC 915
    |||||||
QY 721 caaggttttccacagagatgttgcagcagcagcagcagcagcagcagcagcagcagc 759
    |||||||
Db 916 CAGGCTTTCCACAGATGCTTATGCTGCTGCAATTTGA 954
    |||||||

```

```

RESULT 5
AY040274 1176 bp mRNA linear PRI 01-AUG-2001
LOCUS Homo sapiens apoptosis regulator BCL-2 median form (BCL2) mRNA,
DEFINITION complete cds.
ACCESSION AY040274
VERSION AY040274.1 GI:15072494
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1176)
AUTHORS Montpetit, A., Bolly, G. and Sinnett, D.
TITLE A detailed transcriptional map of the chromosome 12p12 tumor
suppressor locus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1176)
AUTHORS Montpetit, A. and Sinnett, D.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2001) Hemato-Oncology, Hopital Ste-Justine, 3175
Cote-Ste-Catherine, Montreal, QC H3T 1C5, Canada
FEATURES
source
1..1176
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12p12"
1..1176
/gene="BCL2"
8..838

```

gene
CDS

```

RESULT 6
AX207692 984 bp DNA linear PAT 31-AUG-2001
LOCUS AX207692
DEFINITION Sequence 1 from Patent WO0157213.

```


QY	Db	676	GTTTACTCCTGGCCACACACAGGCGAGCCAGGCGAGGGCTTCAAGTCCAAAGACATT	735
QY	541	tttgaactgaagggtctctccctccaagctccaaggccacgctgtatgaattcaagttc	600	
Db	736	TWTGTAACTGAGGGCTCTCTCCCTCCAGCTCCCAAGGCGACGTCGCTAGCTTCAAGTTC	795	
QY	601	aagaagatgaagaagacaaatactagccaaattgttgaactgtctgaatttcsga	660	
Db	796	AAGAAAGATGACAAAGAACAAATACTACCCAAATTTGTAGAGTGTGAATATTACAGA	855	
QY	661	gatcaactggaagaag 678		
Db	856	GATCACTTGGAAGAAG 873		
RESULT	8			
LOCUS	AF281254	1886 bp	mRNA	linear
DEFINITION	Homo sapiens apoptosis regulator BCL-2 long form (BCL2) mRNA,			
VERSION	AF281254			
KEYWORDS	AF281254.1 GI:12584084			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE	1 (bases 1 to 1886)			
JOURNAL	Guo, B., Godzik, A. and Reed, J. C.			
REFERENCE	Bcl-2, a novel pro-apoptotic member of the bcl-2 family			
AUTHORS	J. Biol. Chem. 276 (4), 2780-2785 (2001)			
JOURNAL	2 (bases 1 to 1886)			
REFERENCE	Guo, B., Godzik, A. and Reed, J. C.			
AUTHORS	Direct Submission			
JOURNAL	Submitted (22-JUN-2000) The Burnham Institute, 10901 North Torrey			
FEATURES	Pines Road, La Jolla, CA 92037, USA			
SOURCE	Location/Organisms			
gene	1. 1886			
CDS	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="12"			
	1. 1886			
	/gene="BCL2"			
	196. 1179			
	/gene="BCL2"			
	/note="member of Bcl-2 family"			
	/codon_start=1			
	/product="apoptosis regulator BCL-2 long form"			
	/protein_id="AAG59793.1"			
	/db_xref="GI:12584085"			
	/translation="MSTSGCDLEIRIPDDDLNLTIEFKILAYTRHHVKSYPALFES			
	PKLIRFSLRSGRIGSANSWTEFSPKRSOSSEKAINLKKSSWAFQVVERK			
	EDLSRPAKVSAGSOPRTLEYDSSQMSRCSINPOCIEHEAVDPKVISIANVAEII			
	VTSMPPOATQAGSFKSEIETVTELSQLOSHVPAASSKKDEEDILAKIYELIKTY			
	SGQLERLKKRDMGMHFGDGLSYVEKTTITDVLMDGDPRESEVKAAGFPAALVY			
	DVTAKLEIKIDNPNNRVLFEGTQKYLEKNESPMIIOHGWGEIKILGISHEEVD"			
BASE COUNT	535 a 435 c 450 g 466 t			
ORIGIN				
Query Match	89.1%; Score 676.4; DB 9; Length 1886;			
Best Local Similarity	99.9%; Pred. No. 2.2e-195;			
Matches 677; Conservative	0; Mismatches 1; Indels 0; Gaps 0;			
QY	1	atgtgtagcaccagtggtgtgtgaacctggaagaatccccctagatgatgacctaac	60	
Db	196	ATGTGTAGCACACAGTGCGGTGTGACCTGGAAGAAACCCCTAGATGATGACCTAAC	255	
QY	61	accatagaattcaaatctctgcctctacacacagacagatcatctcttcaagagaccct	120	
Db	256	ACCATGAATTCAAAATCTCGCTCTAGTACACACAGATCATATGCTTCAAGAGACCCCT	315	

OY	121		gctccttccccaagaagctgtcgaaacaagaagttttcaccaaggggacctgggaat	180
Dd	316	GCTTCCTTCCACCAAGACTGTGTAACAAGAAGTTTGTTCCAGAGGGGCGCTGGGAAT		375
OY	181	tgtcgcacaatgatcatgatgacagaggltgtcaitggccttgcaaattcccatacagt	240	
Dd	376	TGTTGCAGCAATATGATCATGTATGACAGAGGTGTTCATGCCCTTGCAGAAATTCCCATAACAGT		435
OY	241	gagaaggccatataacctttgccaaaagaatgtccttttggaaagcatlcttttgaatgtagtg	300	
Dd	436	GAGAAGGCCATAAACCTTAGGCCAAGAAAAAGTCTCCTTTGGAAAAGCATCTTTGGAGTAGTG	495	
OY	301	gaagaagaaatctcgacagagcacgcctgcacaagfctcgcgtccaaaggtccaagaagctlg	360	
Dd	496	GACAGAGCAATATTCGACAGACGACCCTCCCAAAGTCTCTGCTCACGGTCCAAGACCTTG	555	
OY	361	gaataccaagaatlcgcacaagccagcagcagtgctcagaagtgtcctttaagtlgagsgactgc	420	
Dd	556	GAATACCAAAATTTGGCACAGCGCAGAGTGTGTCAGAGTGTCTTTCTPAACTGTGAGCAGTGCG	615	
OY	421	ttggagcataaagtgftggaccccaagaatctaatttcattgtccaaaccgagttaagtaaat	480	
Dd	616	TTGGAGCATPAAGGTGTGTGGACCCCAAAATCATTTTCATTGCCCCAACCGAGTAGCTGAATTT	675	
OY	481	gttatctccgycaccaccaaccaagcgccccagcgagagaggtllcaafccaagaagat	540	
Dd	676	GTTTTACTCCGTGGCCACCACACCAAGCGACGCCAGGCGAGGCTTCAAGTCCAAAGAGATT	735	
OY	541	tttgttaactgaagggtctctctccctccoaagctccaagggccacgtytccgtgtaagcttaagttct	600	
Dd	736	TTTTGTAACTAGAGGTCTCTCTCCTCCAGCTCCCAAGGCCACAGTCCGTAGCTTCAAGTTCT	795	
OY	601	aagaagaataaagaagaacaataactagccaanaattgttgaagctgctgaataatlcaagga	660	
Dd	796	AAGAAGATGATGAGAGACAAATACTTACTACCAAAATTGTGAAGCTGCTGAANAATTTCAGCA	855	
OY	661	gatacagttggaaaaaag 678		
Dd	856	GATCAGTTGCAAGAAGAAG 873		
RESULT LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	9 ACO07537 Homo sapiens Human BAC Library) AC007537 AC007537.3 HTG. human.	190858 bp DNA linear PRI 09-OCT-1999 Homo sapiens 12p12 BAC RPCr11-267j23 (Roswell Park Cancer Institute complete sequence.) GI:4914348		
REFERENCE AUTHORS	Mammalia: Eutheria; Primates; Catarrhini; Homnidae; Homo. Muzny,D., Areson,A.D., Bouck,J.J., Bunae,C., Chen,Z., Ding,Y., Dugan,L., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H., Gorrill,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S., Kondelwski,N., Lau,S., Leal,B., Lee,E., Lichtenage,O., Liu,W., Logen,O., Lu,X., Marondel,I., Martinez,C., Merschke,S., Miller,A., Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,W.D., Rivers,C., Scherer,S.E., Shen,H., Shim,C., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Xiang,A.M., Yang,R., Yu,W., Zhou,X., Kuetherapatti,R., Nelson,D. and Gibbs,R.A.			
TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS	Unpublished 2 (bases 1 to 190858) Worley,K.C. Direct Submission Submitted (13-MAY-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA 3 (bases 1 to 190858)			
REFERENCE AUTHORS	Worley,K.C.			

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

Direct Submission
Submitted (27-MAY-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 190858)
Worley,K.C.
Direct Submission
Submitted (28-MAY-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 190858)
Worley,K.C.
Direct Submission
Submitted (09-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 28, 1999 this sequence version replaced gi:4895156.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
clones with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAR-REPORT-----

Contig length: 190858
Phrap values in estimate: 188824
Average error rate (BCM-Phrap estimate): 0.000239024
Fraction of Phrap values less than 40 : 0.039746
Number of consensus changing edits: 12
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
16186 ataattgaca(c)ccctatctc ataattgaca(a)ccctatctc
26355 taatttga(n)acagcgcaaaa taatttga(a)acagcgcaaaa
35016 tttttttt(n)agcggaagtc tttttttt(g)agcggaagtc
35923 ttattatcat(n)ttttttttt ttattatcat(l)ttttttttt
49346 agcagcgcgga(n)cttcgcagttga agcagcgcgga(g)cttcgcagttga

4963.3		tygccaacat(n)gtgaaacc	tygccaacat(a)gtgaaacc
8051.1	91836	gaagcgagag(n)tgacaacag	gaagcgagag(c)tgacaacag
104278	114732	gaagaaaga(n)accagcttg	gaagaaaga(a)accagcttg
185043	189415	accacttaag(n)tagtagacg	accacttaag(t)tagtagacg
		ctccctccct(n)ccctccctcc	ctccctccct(c)ccctccctcc
		catgtgaa(n)ccgctctca	catgtgaa(c)ccgctctca
		gatcacttga(n)cccaagaagc	gatcacttga(a)cccaagaagc
----- Distribution of Quality < 40 Bases -----			
	10001	9001	8001
	8001	7001	6001
	5001	4001	3001
	2001	1001	0
	*	*	*
	5	10	15
	20	25	30
	35	40	
	Phrap Value Range		

FEATURES	Version: 1.01	xrfgc.	
source	Location/Qualifiers		
	1..190858		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="12p12"		
	/clone="RP111-267023"		
	complement(65..337)		
	/rpt_family="MER7C"		
	338..647		
	/rpt_family="ALUy"		
	complement(648..961)		
	/rpt_family="MER7C"		
	1251..1423		
	/rpt_family="MER5B"		
	1598..1723		
	/rpt_family="MER5A"		
	3236..3327		
	/rpt_family="MIR"		
	complement(3840..4697)		
	/rpt_family="L1MA8"		
	5217..5484		
	/rpt_family="MSTC"		
	5485..5785		
	/rpt_family="ALUSc"		
	5786..5941		
	/rpt_family="MSTC"		
	complement(5976..6074)		
	/rpt_family="HAL1"		
	6170..6191		
	/rpt_family="TTTTG(n)"		
	complement(6284..6456)		
	/rpt_family="MLT1L"		
	complement(6577..6887)		
	/rpt_family="ALUSx"		
	complement(6892..7077)		
	/rpt_family="HAL1"		
	complement(8555..8857)		
	/rpt_family="ALUSx"		
	9274..9448		
	/rpt_family="L1MD"		
	9445..10362		
	/rpt_family="L1MEC"		
	10363..10656		
	/rpt_family="ALUSx"		

```

repeat_region      10657..10870
                    /rpt_family="LIMEC"
repeat_region      10871..11157
                    /rpt_family="AluJo"
repeat_region      11158..11233
                    /rpt_family="LIMEC"
repeat_region      11234..11520
                    /rpt_family="AluSx"
repeat_region      11521..11938
                    /rpt_family="LIMEC"
repeat_region      complement(12005..12070)
                    /rpt_family="AluS"
repeat_region      12071..12253
                    /rpt_family="LIMD1"
repeat_region      complement(12254..12648)
                    /rpt_family="MLT1A1"
repeat_region      12649..14923
                    /rpt_family="LIMD1"
repeat_region      14924..15235
                    /rpt_family="AluY"
repeat_region      15236..15588
                    /rpt_family="LIMD1"
repeat_region      complement(15646..15821)
                    /rpt_family="MLT1H"
repeat_region      15838..16079

```

```

Query Match      57.4% Score 435.8 DB 9 Length 190858;
Best Local Similarity 97.4% Pred. No. 1.1e-121;
Matches 443; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

OY 1 atgtgtagaccagctgggtgtgacctggaagaatccccctagatgatgaactaac 60
DB 48512 AGCTGAGCAGCAGTGGGTGACCTGGAAGAAATCCCTAGATGATGATCAAC 48571
OY 61 accatagaatcaaatccctgcctactacccagacatcatgttccaagaccct 120
DB 48572 ACCATAGAATCAAAATCCCTGCTACTACACGACATCATGCTTCAAGACACCCCT 48631
OY 121 gctctctccaccaagctgctgagaagaagttgtcccaagggcctggggaat 180
DB 48632 GCCTCTTCCACCAAGCTGCTGAGAACAAAGTTGTCCCAAGGGGCTGGGGAA 48691
OY 181 tcttcagcaaatgagtcagacagagtgatcagcttgagcaaatcccaatcagt 240
DB 48692 TCTTCAGCAAAATGATGATGAGACAGAGGTGATGCTGCAAGAAATCCCAATCAGT 48751
OY 241 gagaagggccataacctgtgcaagaagaatctcttcttgagaagcatcttcttgagtagtg 300
DB 48752 GAGAAAGGCCATAAAGCTTGCAAGAAAGTCTTCTTGAAAGCATCTTCTTGAGTAGTG 48811
OY 301 gagaagaagatctgcagacagcctgcagagttctctgctcagagtcagaagcgttg 360
DB 48812 GAGAAAGGAATTCGCGAGAGCAGCCTGCAAGGCTCTGCTCAAGGTCAAAGAGAGCTTG 48871
OY 361 gaataccaagatctgcagacagcagctgtgctcaggttcttcttaacgtgagcagtg 420
DB 48872 GAATACCAAGATTCGCGAGAGCAGAGGTGCTCAAGGTCCTTCTTAAGCTGAGAGCTG 48931
OY 421 ttggaagcatgaagctgtggaaccccaagtcatttc 455
DB 48932 TTGGAAGCATGAAGGTAGGATCTGGAATTTCTTTC 48966

```

```

RESULT 10
AC022222/c AC022222 194143 bp DNA linear HTG 15-OCT-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-52513, WORKING DRAFT
SEQUENCE 26 unordered pieces.
AC022222
AC022222
AC022222.34 GI:14547295
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens

```

REFERENCE

AUTHORS

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 194143)

Muzny,D.M., Adams,C., Adio-Obuola,B., All-ouman,F.R., Allen,C., Alstbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbria,J., Benton,J., Bimage,K., Blankenburg,K., Bonini,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Diaper,H., Dugan-Rocha,S., Durbin,K.J., Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B., Homsif,F., Howard,S., Huber,J., Huliy,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,J.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Louisedge,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Plickens,R., Prims,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuboan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoshchari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleceyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and GIBBS,R.

Direct Submission
Unpublished
2 (bases 1 to 194143)

Worley,K.C.
Direct Submission
Submitted (27-JAN-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:13957543.

COMMENT

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMBL
Center clone name: RP11-52513
----- Summary Statistics
Sequencing vector: Plasmid; M77789
Chemistry: Dye-terminator Big Dye; 93% of reads
Chemistry: Dye-terminator Big Dye; 93% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 185441 bases at least Q40
Consensus quality: 200072 bases at least Q30
Consensus quality: 208935 bases at least Q20
Estimated insert size: 199196; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 3.3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

1	21477	21476:	contig of 21476 bp in length
*	21577	21576:	gap of unknown length
*	41551	41450:	contig of 19874 bp in length
*	41551	41550:	gap of unknown length
*	55299	55299:	contig of 13748 bp in length
*	55299	55399:	gap of unknown length
*	66765	66764:	contig of 11366 bp in length
*	66765	66864:	gap of unknown length
*	81568	81567:	contig of 14703 bp in length
*	81568	81667:	gap of unknown length
*	94374	94373:	contig of 12706 bp in length
*	94474	104139:	contig of 9666 bp in length
*	104140	104123:	gap of unknown length
*	104240	114455:	contig of 10214 bp in length
*	114554	114553:	gap of unknown length
*	114554	122925:	contig of 8372 bp in length
*	122926	123025:	gap of unknown length
*	123026	127231:	contig of 4206 bp in length
*	127232	127331:	gap of unknown length
*	127332	134754:	contig of 7423 bp in length
*	134755	134755:	gap of unknown length
*	134855	140454:	contig of 5600 bp in length
*	140555	140554:	gap of unknown length
*	140555	147533:	contig of 6581 bp in length
*	147536	147535:	gap of unknown length
*	147636	153065:	contig of 5434 bp in length
*	153070	153169:	gap of unknown length
*	153170	159234:	contig of 6065 bp in length
*	159335	159334:	gap of unknown length
*	159335	164821:	contig of 5493 bp in length
*	164828	164921:	gap of unknown length
*	164828	169721:	contig of 4794 bp in length
*	169722	169821:	gap of unknown length
*	169822	172655:	contig of 2744 bp in length
*	172656	172655:	gap of unknown length
*	172656	175115:	contig of 2450 bp in length
*	175116	175215:	gap of unknown length
*	175116	177736:	contig of 2343 bp in length
*	177736	177858:	gap of unknown length
*	177859	180950:	contig of 3092 bp in length
*	180951	181051:	gap of unknown length
*	181051	184037:	contig of 2987 bp in length
*	184038	184137:	gap of unknown length
*	184138	184793:	contig of 3358 bp in length
*	184796	187595:	gap of unknown length
*	187596	189691:	contig of 2096 bp in length
*	189692	189791:	gap of unknown length
*	189792	191812:	contig of 2021 bp in length
*	191813	191912:	gap of unknown length
*	191913	194143:	contig of 2231 bp in length

	a	c	g	t	others
BASE COUNT	54237	42567	41198	53596	2545
ORIGIN					

03

QY	1	atgtttagacacccgttgggtgttgaccttggaaataatcccccttagatgatgatgaac	60
Db	177208	ANGTGTAGACACCGTGGGTGTGACCTTGGAAATAATCCCCCTTGATGATGATGACCTTAAC	177149
QY	61	accatagaattcaaatcctgcctactaaccagaacatcatgcttcaagacacct	120
Db	177148	ACCAATAGAAATTCCTAAATCCCTGGCTTACTACCCAGACATCATGTCTTCAAGAGACACCCCT	1770883
QY	121	gctctctctcaccaaagctgctctgagaagaagaagttgttcccaaggagccttgggaat	180
Db	177088	GCTCTCTTCTCACCAAGCTGCTGTAAGACAAAGAGTTGTCTCCAAAGAGGGGCTGGGGAAAT	1770293
QY	181	tgttcagcaaatgagtcatgtagacagaggtgtcatgaccttgcagaataatccccatcagt	240
Db	177028	TGTTGACGAATGATGATCATGTGAGCAGAGAGTGTCATGAGCCTTGAGAAATCCCAATCAGT	176965
QY	241	gaggaaggccataaactcttgccaagaataaagtctctcttggaaagcattcttggagtatg	300
Db	176968	GAGAAAGCCATTAACCTTTGGCAGAAAAAAGCTCTTTGGAAGCAATCTTTGGAGATG	1769083
QY	301	gagaaaggaaatctgcagacagcacgcctgccaaggtctctgcacaggttcaaaagacgttg	360
Db	176908	GAGAAAGGAATTCGCGACAGACACCCCTTGCCAAAGTCTTGCTCAAGGTCTAAAGAGACGTG	1768493
QY	361	gaataaccaagatctgcacagccagccagctgtgtccagtgcttcttaacgttgcagtcagtc	420
Db	176848	GAAATCCCAAGATTCGCGACACACCCAGCAGTGTGCTCAAGTTCATTAACTGGAGAGCAGTGC	1767883
QY	421	ttgagacatgaacgtgtgagccccaagaatcttc	455
Db	176788	TTGAGACATGAAGTGAAGCATCTGGGATTTCTTTTC	176754

RESULT	11
AC091816	
LOCUS	
DEFINITION	Homo sapiens chromosome 12p clone RP11-525I3, WORKING DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION	AC091816
VERSION	AC091816.7 GI:15625994
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 200499)
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.
1 (bases 1 to 200499)

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alstbrooks, S.L., Amaralunga, H.C., Ate, J.R., Banks, T., Barbata, J., Benton, J., Bimarg, K., Blankenburg, K., Bonini, D., Boucek, J., Bowler, S., Brieleva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, J., Cleveland, C.D., Cox, C., Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Degen, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J., Hernandez, O., Hodson, A., Hogue, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolyvet, S., Koudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Koryah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Lounsbeg, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, U., Maheshwari, M., Mapua, P., Martin, R., Matindale, A., Martinez, E., Messey, E., Mawhinney, E., Melcoid, M.P., Meador, M., Melzer, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morjan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,

BASE COUNT 272 a 246 c 271 g 197 t 1 others
 ORIGIN

Query Match 36.2% Score 274.4: DB 6; Length 987;
 Best Local Similarity 67.7% Pred. No. 1.5e-72;
 Matches 469; Conservative 0; Mismatches 197; Indels 27; Gaps 5;

```

QY 1 atgcttagcaccagctgggtgtgacctggaagaatccccctagatgatgtgacctaac 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 ATGCGACGACACAGCTGTATGACCTGGAAAGACATTCCTCGAGATGTATGATCCAAAC 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 accatagaattcaaaatccctgcctactacacacagacatcatgtcttcaagagcaccct 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 AGCATAGATTCAAAATCTGGCTTCTTACGCCAGACACACATGCTTCAAGAACACCCCG 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 gctctctctacaaaactctggaagaagaagtcttcccaaggaggcctggggaat 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 GCTCTCTCTGCCCCAACCTCTCCAGAACAGAGCTGTCTCCAGAAAGCCCTGGGACT 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 tttcagaacaaatgacatgacagagagtgatgacgtctgcaaaatcccaatccagt 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 TGTGT---CAACTGATTCCTGGACACAGATATCATTCCTTGCAGAGTTCCCTCCAGC 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 241 gagaagaccataaactctggcaagaagaagtcttcttgaagaactcttggagtagtg 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 GAAAGAGAACATCACCTTGGCGAAGAAAGTCTTCTTGGAGAACACTCTCAGGCTGCC 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 301 gaga---aggaagattgcagagacagcgcctgcgaaggtcttgcctaaaggagcg 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 GAGAGAGAGGAGAGGCTGCGCTGCCCAAGAGATCGAGCTCAGGCTCCTCAGAGGC 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 358 ttggaataccagaattcgc-----cacagccagcagtgctgcaaggtcttctt 405
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 358 CCCCTCCCGGTAGAGAGCGGACAGTGGCTTCCACACACACACACAGCTGCTCTAGC 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 406 aagctggagcagctctggaagcagatgaagctgtggaaccccaagtcattccattgccaac 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 406 AGCTGTGAGGAGG---CCCTGAGAGTGAAGTTTGATTCCTCAAGTGGCTGTATGGCAAC 476
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 466 cgaagtagctgaattgttattctctgcccacacacacacagcagcccaagcagaagagcttc 525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 477 AGAGTGTGCTGAATGTGTACTCTCTGSCACACACAGATGTATCCACAGCCAGGAGGGA 536
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 526 aagtcacaagagattttgttaactagagggtctctctctccacagcagcagcagctcct 585
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 537 AGGCAGCTCAAGAGAGGCTCTCGGAGATTTTGTACTTCAGGTTTGAAGACTTGGGA 596
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 586 gtgactccaagttctaaagaagaagaacaatactagccaataattgttgagctg 645
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 597 CTC-----TAAGAAATTAAGATGTGTGAAGCCAAATTAAGCAAGATTGTGAGCTG 648
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 646 ctgaatatctcagagagatcagttggaagaagaag 678
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 649 CTGAATCTCTGGGGGATCACTTGGGAAGAGAG 681
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

RESULT 13
 AC097248/c 149350 bp DNA linear HNG 20-Dec-2001
 LOCUS Rattus norvegicus clone CH230-177C11, *** SEQUENCING IN PROGRESS
 DEFINITION *** 64 unordered pieces.
 AC097248
 VERSION AC097248.3 GI:17973829
 KEYWORDS HNG; HNGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 149350)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alstbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,

Benton,J., Binage,K., Blankenburg,K., Bonjin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Denny-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodson,A., Hogue,M., Holloway,C.,
 Hollins,B., Homsli,F., Howard,S., Huber,J., Hultys,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,D., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W.,
 Louisedge,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
 Ogunb,M., Okunou,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
 Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
 Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Taber,P., Tameris,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalon,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Wallington,S., Williams,G., Williams,A., Wleczek,R., Wooden,S.,
 Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 149350)
 Worley,K.C.
 Direct Submission
 Submitted (13-Oct-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Dec 20, 2001 this sequence version replaced gi:17064657.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GINL
 Center clone name: CH230-177C11
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329first call to
 findphraplist
 Consensus quality: 114307 bases at least Q40
 Consensus quality: 121845 bases at least Q30
 Consensus quality: 128763 bases at least Q20
 Estimated insert size: 114370; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agrose-fp estimation
 Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 64 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will


```

*****
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
5539: contig of 5539 bp in length
5540 5539: gap of unknown length
5540 10594: contig of 4955 bp in length
10595 10594: gap of unknown length
10595 16555: contig of 5871 bp in length
16556 16555: gap of unknown length
16556 21899: contig of 5234 bp in length
21900 21899: gap of unknown length
26847 21900: contig of 4848 bp in length
26848 26847: gap of unknown length
30996 26848: contig of 4048 bp in length
31096 30996: gap of unknown length
35906 31096: contig of 4810 bp in length
36006 35906: gap of unknown length
40022 36006: contig of 4016 bp in length
40122 40022: gap of unknown length
45189 40122: contig of 5067 bp in length
45289 45189: gap of unknown length
49459 45289: contig of 4171 bp in length
49559 49459: gap of unknown length
51796 49559: contig of 2237 bp in length
51897 51796: gap of unknown length
54862 51897: contig of 2966 bp in length
54863 54862: gap of unknown length
54963 54863: contig of 3191 bp in length
58154 54963: gap of unknown length
58254 58154: contig of 2490 bp in length
60744 58254: gap of unknown length
60844 60744: contig of 2832 bp in length
63776 60844: gap of unknown length
63776 63776: gap of unknown length
66399 63776: contig of 2624 bp in length
66400 66399: gap of unknown length
69667 66400: contig of 3368 bp in length
69668 69667: gap of unknown length
72345 69668: contig of 2378 bp in length
72346 72345: gap of unknown length
73890 72346: gap of unknown length
73891 73890: contig of 1445 bp in length
76130 73891: gap of unknown length
76131 76130: contig of 2140 bp in length
77990 76131: gap of unknown length
77991 77990: contig of 1760 bp in length
78091 77991: gap of unknown length
78091 78091: contig of 1776 bp in length
79667 78091: gap of unknown length
79667 82089: contig of 2122 bp in length
82089 79667: gap of unknown length
82189 82089: contig of 1968 bp in length
84157 82189: gap of unknown length
84257 84157: gap of unknown length
86357 84257: contig of 2101 bp in length
86358 86357: gap of unknown length
86458 86358: gap of unknown length
87963 86458: contig of 1506 bp in length
88063 87963: gap of unknown length
89282 88063: contig of 1219 bp in length
89382 89282: gap of unknown length
89383 89382: contig of 1665 bp in length
91047 89383: gap of unknown length
91048 91047: gap of unknown length
93054 91048: contig of 1907 bp in length
93154 93054: gap of unknown length
93155 93154: contig of 2023 bp in length
95277 93155: gap of unknown length
95278 95277: contig of 1633 bp in length
96910 95278: gap of unknown length
97010 96910: contig of 1633 bp in length

```

```

* 97011 98848: contig of 1838 bp in length
* 98849 98848: gap of unknown length
* 98949 100455: contig of 1507 bp in length
* 100456 100555: gap of unknown length
* 100556 102754: contig of 2199 bp in length
* 102755 102854: gap of unknown length
* 102855 104635: contig of 1781 bp in length
* 104636 104735: gap of unknown length
* 104736 106420: contig of 1665 bp in length
* 106421 106520: gap of unknown length
* 106521 107762: contig of 1242 bp in length
* 107763 107762: gap of unknown length
* 107863 109530: contig of 1668 bp in length
* 109531 109630: gap of unknown length
* 109631 111109: contig of 1479 bp in length
* 111110 111209: gap of unknown length
* 111210 113098: contig of 1889 bp in length
* 113099 113198: gap of unknown length
* 113199 114273: contig of 1075 bp in length
* 114274 114373: gap of unknown length
* 114374 116703: contig of 2330 bp in length
* 116704 116803: gap of unknown length
* 116804 117940: contig of 1137 bp in length
* 117941 118040: gap of unknown length
* 118041 120245: contig of 2205 bp in length
* 120246 120345: gap of unknown length
* 120346 121494: contig of 1149 bp in length
* 121495 121594: gap of unknown length
* 121595 123183: contig of 1589 bp in length
* 123184 123283: gap of unknown length
* 123284 124580: contig of 1297 bp in length
* 124581 124680: gap of unknown length
* 124681 126245: contig of 1565 bp in length
* 126246 126345: gap of unknown length
* 126346 127733: contig of 1388 bp in length
* 127734 127833: gap of unknown length
* 127834 129130: contig of 1297 bp in length
* 129131 129230: gap of unknown length
* 129231 130800: contig of 1570 bp in length
* 130801 130900: gap of unknown length
* 130901 132501: contig of 1601 bp in length
* 132502 132601: gap of unknown length
* 132602 134552: contig of 1951 bp in length
* 134553 134652: gap of unknown length
* 134653 135941: contig of 1289 bp in length
* 135942 136041: gap of unknown length
* 136042 137656: contig of 1615 bp in length
* 137657 137756: gap of unknown length
* 137757 138840: contig of 1084 bp in length
* 138841 138940: gap of unknown length

```

Query Match 8.2% Score 62.4; DB 2; Length 149350;
 Best Local Similarity 57.0% Pred. No. 1.4e-07;
 Matches 114; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

QY 412 gacgagtcctggagcagcgtgtgagcccaagcattccattgccaacggatga 471
||||| ||| ||||||| ||||||| ||||||| ||| |||
Db 56394 GAGCAGTCTTCTCGTTTCACCTGCGATTCGCAAGTTCGTTATGCAACAGAGTG 56453
||||| ||| ||||||| ||||||| ||||||| ||| |||
QY 472 gctgaattgttattctctgcgcaccacacagcagccagcagcagcagcagcagtc 531
||||| ||| ||||||| ||||||| ||||||| ||| |||
Db 56454 GCTGAATGATTTACTCTCGGCGACACCGAGATGACGAGCCAGGAGGAGGCAAG 56513
||||| ||| ||||||| ||||||| ||||||| ||| |||
QY 532 aaagagatttgaactgaagtgctctccagcagcagcagcagcagcagcagcagtc 591
||||| ||| ||||||| ||||||| ||||||| ||| |||
Db 56514 TTCGAAGAGAGTGTCGCAAGATTCGACATTGCAATGCAAGGCGCTCAACGAGACC 56573
||||| ||| ||||||| ||||||| ||||||| ||| |||
QY 592 tcaagttcagaagaagatga 611
||||| ||| ||||||| |||
Db 56574 TGCACCTCGAAGAAAGGTGA 56593
||||| ||| ||||||| |||

```

Search completed: June 20, 2002, 05:11:00
Job time: 9460 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2002, 16:39:44 ; Search time 46.53 Seconds
(without alignments)
936.916 Million cell updates/sec

Title: US-09-771-961-4
Perfect score: 1316
Sequence: 1 MCTSGCDLEETPLDDDDLTIEFKILAYTRHHVFKSTPALFSPKILRTSLSGRGLGN 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 562222
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1316	100.0	252	4	Q9BZr7 homo sapien
2	1189.5	90.4	327	4	Q9BZr8 homo sapien
3	1181	89.7	276	4	Q9BQR5 homo sapien
4	651	49.5	328	11	Q9CPT0 mus musculu
5	650	49.4	328	11	Q9D3W3 mus musculu
6	97	7.4	450	3	Q13722 schizosacch
7	95.5	7.3	1253	11	Q9EQS8 mus musculu
8	95	7.2	1083	10	Q48B39 arabidopsis
9	93	7.1	980	4	Q9HCE4 homo sapien
10	91.5	7.0	960	6	Q18965 bos taurus
11	91.5	7.0	987	6	Q18966 bos taurus
12	89.5	6.8	647	16	Q9WZEO thermotoga
13	89.5	6.8	696	5	Q9VK07 drosophila
14	89	6.8	1252	11	Q9EQS9 mus musculu
15	89	6.8	1252	11	Q9J1L1 mus musculu
16	89	6.8	1262	5	Q9N911 leishmania

17	89	6.8	1293	10	Q9AXD6 zea mays (m
18	88.5	6.7	250	11	Q9CR18 mus musculu
19	88	6.7	638	11	Q99KW6 mus musculu
20	88	6.7	4202	11	Q91V63 mus musculu
21	88	6.7	4731	11	Q91XP8 mus musculu
22	87.5	6.6	287	3	Q43030 schizosacch
23	87.5	6.6	791	11	Q9C2X2 mus musculu
24	87.5	6.6	3115	5	Q95Y72 leishmania
25	87	6.6	1660	10	Q9SIR1 arabidopsis
26	85.5	6.5	464	16	P74199 synechocyst
27	85.5	6.5	889	10	Q9M1M3 arabidopsis
28	85.5	6.5	1311	5	Q9V899 drosophila
29	85	6.5	1080	10	Q9V3028 arabidopsis
30	84	6.4	336	10	Q93364 brassica ol
31	84	6.4	494	15	Q9QC00 human immun
32	84	6.4	724	15	Q83009 lymphoprol
33	83.5	6.3	497	15	Q901A1 human immun
34	83.5	6.3	1284	5	Q23974 drosophila
35	83	6.3	332	16	Q98Q48 mycoplasma
36	83	6.3	489	11	Q9D4A0 mus musculu
37	83	6.3	489	11	Q91XT6 mus musculu
38	83	6.3	1190	16	Q9C1E1 lactococcus
39	83	6.3	1193	16	Q9HZE5 pseudomonas
40	83	6.3	1264	4	Q96E77 homo sapien
41	83	6.3	1447	10	Q64473 arabidopsis
42	83	6.3	3477	11	Q99MH8 mus musculu
43	83	6.3	4202	11	Q91XP9 mus musculu
44	83	6.3	4731	11	Q91XG0 mus musculu
45	82.5	6.3	258	10	Q9XJ56 daucus caro

ALIGNMENTS

RESULT	ID	Query Match	Score	1316:	DB 4:	Length	252:
1	Q9BZr7	100.0%	1316	100.0%	Pred. No. 2.1e-115;	Matches 252;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
AC	Q9BZr7	PRELIMINARY;	PRT;	252 AA.			
DT	01-JUN-2001 (Tremblrel. 17, Created)						
DT	01-JUN-2001 (Tremblrel. 17, Last sequence update)						
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)						
DE	APOPTOSIS REGULATOR BCL-6 SHORT FORM.						
GN	BCL6.						
OS	Homo sapiens (Human).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.						
OX	NCBI_TaxID-9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RX	MEDLINE-21264734; PubMed-11054413;						
RA	Guo B., Godzik A., Reed J.C.;						
RT	"Bcl-6, a novel pro-apoptotic member of the Bcl-2 family.";						
RL	J. Biol. Chem. 276:2780-2785(2001).						
DR	EMBL, AF281255; AAG59794.1;						
SO	SEQUENCE 252 AA; 28089 MW; 87D2E5123EFCB9E4 CRC64;						

Query Match	Score	1316:	DB 4:	Length	252:
Best Local Similarity	100.0%	Pred. No. 2.1e-115;	Matches 252;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	MCTSGCDLEETPLDDDDLTIEFKILAYTRHHVFKSTPALFSPKILRTSLSGRGLGN 60			
DB	1	MCTSGCDLEETPLDDDDLTIEFKILAYTRHHVFKSTPALFSPKILRTSLSGRGLGN 60			
QY	61	CSANESWTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVKEKDSQSTPAKVSAGGRTL 120			
DB	61	CSANESWTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVKEKDSQSTPAKVSAGGRTL 120			
QY	121	EXODSHSQMSRCLSNVEOELEHADVPKVISTANRYAEIYSSMPPOATGAGFKREI 180			
DB	121	EXODSHSQMSRCLSNVEOELEHADVPKVISTANRYAEIYSSMPPOATGAGFKREI 180			

QY	QY	QY	QY
181	181	181	181
FVTEGSLFQIQGHVPAVSSSKDEEOLIAKIVELTKYSSDQLERKOTAPILPIVDTSI	FVTEGSLFQIQGHVPAVSSSKDEEOLIAKIVELTKYSSDQLERKOTAPILPIVDTSI	FVTEGSLFQIQGHVPAVSSSKDEEOLIAKIVELTKYSSDQLERKOTAPILPIVDTSI	FVTEGSLFQIQGHVPAVSSSKDEEOLIAKIVELTKYSSDQLERKOTAPILPIVDTSI
241	241	241	241
QGFPODGLMAGI 252	QGFPODGLMAGI 252	QGFPODGLMAGI 252	QGFPODGLMAGI 252
241	241	241	241
QGFPODGLMAGI 252	QGFPODGLMAGI 252	QGFPODGLMAGI 252	QGFPODGLMAGI 252

RESULT	2	
Q9BZR8		
ID	Q9BZR8	PRELIMINARY;
000000		PRT; 327 AA.

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE APOPTOSIS REGULATOR BCL-G LONG FORM.
 GN BCLG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;

RP SEQUENCE FROM N.A. PubMed=11054413;
RX MEDLINE=21264734; Godzik A., Reed J.C.;
RA Guo B., a novel pro-apoptotic member of the Bcl-2 family
RT "Bcl-2, a novel pro-apoptotic member of the Bcl-2 family
RL J. Biol. Chem. 276:2780-2785(2001).
DR EMBL; AF281254; AAG59793.1; -;
DR InterPro; IPR002475; Bcl2_family.
DR InterPro; IPR000712; Bcl2_2.
DR Pfam; PF00452; Bcl-2; 1.
DR SMART; SMO0337; BCL_1.
DR PROSITE; PS50062; BCL2_FAMILY; 1.
SQ SEQUENCE 37 AA; 36598 MW; 15D543BA91BFB3EA CRC64

Query Match	90.4%	Score 1189.5;	DB 4;	Length 327;
Best Local Similarity	93.5%;	Pred. No. 2.2e-103;		
Matches 232; Conservative	2;	Mismatches 7;	Indels 7;	Gaps 12

QY	1	MCSHSGDLEIFPLDDDDLMNTEKILAYATRHNVFSTPALSPKLRRSLSORGLGN	60
Db	1	MCSHSGDLEIFPLDDDDLMNTEKILAYATRHNVFSTPALSPKLRRSLSORGLGN	60
QY	61	CSANESWTEVSWPCRNSSOSEKAIINLGKKSSWKAFFGVVEKEDSOSTPAKVASNOGRTL	120
Db	61	CSANESWTEVSWPCRNSSOSEKAIINLGKKSSWKAFFGVVEKEDSOSTPAKVASNOGRTL	120
QY	121	EYODSHSOOMSRLCSINVEOCLHEHAVDPKVIYSIANRVAEIVYSWPPQAOAGGFSKKEI	180
Db	121	EYODSHSOOMSRLCSINVEOCLHEHAVDPKVIYSIANRVAEIVYSWPPQAOAGGFSKKEI	180
QY	181	FVTEGLSFJOLGHPVAVSSSKKDEEOILAKIVELLKYSGDQLEKRDPAEIPILVDTST	240
Db	181	FVTEGLSFJOLGHPVAVSSSKKDEEOILAKIVELLKYSGDQLEKRLTK-----DKAL	233
QY	241	QGFPDGL 248	
Db	234	MGHFQDGL 241	

RESULT	3	
Q96QR5		
ID	Q96QR5	PRELIMINARY;
		PRT;
		276 AA.

DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE APOPTOSIS REGULATOR BCL-G MEDIAN FORM.
GN BCL6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Montpetit A., Bolly G., Smet D.,
RT "A detailed transcriptional map of the chromosome 12p12 tumor
suppressor locus".
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
SO EMBL: AY040274; MAK72109.1; -
DR SEQUENCE 276 Aa; 30948 Mw; 81559A7190F59BE CMC64;

Query Match	89.78;	Score 1181;	DB 4;	Length 276;
Best local Similarity	99.68;	Pred. No. 1.1e-102;		
Matches 226;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MCSTSGCDLEETPLDDDDDLNTEKILAYATYRHVHFSTALPSEPKLLRFRSLSORGLGN	60
Db	1	MCSTSGCDLEETPLDDDDDLNTEKILAYATYRHVHFSTALPSPKLLRFRSLSORGLGN	60
QY	61	CSANESWTEYSWPCRNSSOSEKALNLRGKKSSWMAFEFVYEKEDSOSTPAKVSNOGRTL	120
Db	61	CSANESWTEYSWPCRNSSOSEKALNLRGKKSSWMAFEFVYEKEDSOSTPAKVSNOGRTL	120
QY	121	EYDQSHSQMSQRCLSNWEQCLLEHNAVDPKYTISIANRAEAIYISMPPOQAIQAGFSKKEI	180
Db	121	EYDQSHSQMSQRCLSNWEQCLLEHNAVDPKYTISIANRAEAIYISMPPOQAIQAGFSKKEI	180
QY	181	FVTEGLSFOLOGHVPVASSSKKDEEEOILAKTIVELLKYSGDQLERKD	227
Db	181	FVTEGLSFOLOGHVPVASSSKKDEEEOILAKTIVELLKYSGDQLERKE	227

RESULT	4	
Q9CPT0		
ID	Q9CPT0	PRELIMINARY;
328	328	PRT;
AA	AA	

DT 01-JUN-2001 (TREMBLERel. 17, Created)
 DT 01-JUN-2001 (TREMBLERel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLERel. 17, Last annotation update)
 DE 9030625M01R1K PROTEIN (4933405K19R1K PROTEIN).
 GN 9030625M01R1K OR 4933405K19R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Murinae; Mus.
 NX NCBI_faxID=10090;

RC STEAIN157BL/6J; TISSUE=COLON, STOMACH, AND TESTIS;
RX MEDLINE=21085660; PUBMED=11217651;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Izawa A., Fukunishi Y., Konno H., Ando S., Fukuda S.,
RA Aikawa K., Hara M., Nishii K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Caasteleland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikrdo I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustlinich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaets P.,
RA Nordone P., Rong B., Ringwald C., Rodriguez I., Sakamoto N.,
RA Saasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoo-Oka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshnaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK018579; BAB31290.1; -
DR EMBL; AK008682; BAB25830.1; -
DR EMBL; AK016670; BAB30370.1; -


```

ID 09HCE4 PRELIMINARY; PRT; 980 AA.
AC 09HCE4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE KIAA1628 PROTEIN (FRAGMENT).
GN KIAA1628
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20450683; Pubmed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT Prediction of the coding sequences of unidentified human genes.
RT XVII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.
EMBL: AB046848; BAB13454.1;
DR InterPro: IPR003962; FNIL_repeat.
DR InterPro: IPR003961; FN_IL.
DR InterPro: IPR003598; Iq_c2.
DR InterPro: IPR003600; Iq_like.
DR pfam: PF00041; fn3; 5.
DR pfam: PF00047; Iq; 1.
DR PRINTS: PRO0014; FNTYPEIII.
DR SMART: SM00060; FN3; 5.
DR SMART: SM00409; Iq; 1.
DR SMART: SM00408; IqC2; 1.
DR SMART: SM00410; Iq_like; 1.
KW Immunoglobulin domain; Repeat.
FT NON_TER
SQ SEQUENCE 980 AA; 105361 MW; A0636C95D4E951E CRC64;

Query Match 7.1%; Score 93; DB 4; Length 980;
Best Local Similarity 24.4%; Pred. No. 6.6;
Matches 47; Conservative 25; Mismatches 83; Indels 38; Gaps 8;

QY 8 DLEIPLDDDDLTN-IEKILAYIRHVFKESTPALFS-----PKLLRRLSLQRLGNC 61
DB 214 DTLEQLQVRLDEPNTDYEVYVAVSQLGASRTSPALVHTLDPVPSAAPOLSL- 266
QY 62 SANESWTEVSW-PCRNSSSEKATILGKKSSWKAFFGVKEKDSSTPAK-----V 112
DB 267 SPNPSDIRVAVLPLPPSLNSGOVY-----KKTLEYGLGKEQIPSTEVNGNETOLMLN 319
QY 113 SAOQRTLEYQDSHS-----QQMSRLSNVEQCLEHEAVDPKVISIANRYAEIYYSW 164
DB 320 SLDPNKYVRIRISAGTAAGFAPSCWMHHRTPSMNHQSHVPPAPALPKVQAKMESLVYSW 379
QY 165 -PPQATQAGGFK 176
DB 380 QPPPHPTQISGYK 392

RESULT 10
O18965 PRELIMINARY; PRT; 960 AA.
AC O18965;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE EAG CHANNEL.
GN EAG1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98190099; Pubmed=9524140;
RA Frings S., Bruell N., Dzeja C., Angele A., Hagen V., Kaupp U.B.,
RA Baumann A.;
RT "Characterization of ether-a-go-go channels present in photoreceptors
RT reveals similarity to Ix, a K current in rod inner segments.";
RL J. Gen. Physiol. 111:583-599(1998).
DR EMBL: Y13430; CAA73842.1;
DR InterPro: IPR000636; Cation_chan_non_11g.
DR InterPro: IPR001622; Channel_pore_K.
DR InterPro: IPR000595; cNMP_binding.
DR InterPro: IPR001610; PAC.
DR InterPro: IPR000700; PAS-assoc_C.
DR pfam: PF00027; cNMP_binding; 1.
DR pfam: PF00520; ion_trans; 1.
DR pfam: PF00785; PAC; 1.
DR SMART: SM00100; cNMP; 1.
DR SMART: SM00086; PAC; 1.
DR SMART: SM00091; PAS; 1.
DR PROSITE: PS0042; cNMP_BINDING_3; 1.
SQ SEQUENCE 960 AA; 108092 MW; 86fCD3303A3E3B08 CRC64;

Query Match 7.0%; Score 91.5; DB 6; Length 960;
Best Local Similarity 21.3%; Pred. No. 8.9;
Matches 49; Conservative 40; Mismatches 80; Indels 61; Gaps 10;

QY 15 DDDDLNTIEFKILAYIRHVFKEST-----PALFSPKLL--RTNSLSORG 57
DB 733 DDDDLVERKYSVLTESHSHGLAKASVTVRESPATPVAFPAAPAGIDHRLQAPGAEG 792
QY 58 IG-----NCSANESWTEVSWPCRNSSSEKATILGKKSSWKAFFGVKEKDSSTPAK 111
DB 793 LEPKAGGADCAKRGKGNAREFDACGQAEWMSK-----VSKAESMTLPER 836
QY 112 VSAOQRTLEYQDSHSQOMSRC--LSNVEQCLEHEAVDPKVISIANRYAEIYVS-WP-P 166
DB 837 TKAAGATLAKNTDSCSGITKSDLRLDNVGEARSPODRSP-----TLAEVKSIFYIP 889
QY 167 PQATQAGGKSKEIFYTEGLSFQLOGHVPAVSSSKKDEEOITAKIYELL 216
DB 890 EQTLQAAVLEVKH-----ELKEDIK-ALSTKMTSIEKQLSEILRIL 929

RESULT 11
O18966 PRELIMINARY; PRT; 987 AA.
AC O18966;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE EAG CHANNEL.
GN EAG2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98190099; Pubmed=9524140;
RA Frings S., Bruell N., Dzeja C., Angele A., Hagen V., Kaupp U.B.,
RA Baumann A.;
RT "Characterization of ether-a-go-go channels present in photoreceptors
RT reveals similarity to Ix, a K current in rod inner segments.";
RL J. Gen. Physiol. 111:583-599(1998).
DR EMBL: Y13431; CAA73843.1;
DR InterPro: IPR000636; Cation_chan_non_11g.
DR InterPro: IPR001622; Channel_pore_K.
DR InterPro: IPR000595; cNMP_binding.
DR InterPro: IPR001610; PAC.

```



```

DR FlyBase: FBgn0032499; CG6116.
SQ SEQUENCE 696 AA; 80690 MW; 26F0DA367CF8D0CD CAC64;

Query Match
Best Local Similarity 21.0%; Score 89.5; DB 5; Length 696;
Matches 48; Conservative 29; Mismatches 77; Indels 75; Gaps 11.

OY 6 GCDLEIP-----LDDDLNTIEKTIATYTRHHVFSTPALSPKILIRSLSQIG 59
DB 28 GGNIESWHEKSLIEDDDP-----DVLlyYTL-HTDASEAFYTSKPLPQRHQQKMAE 80
OY 60 NCSANSESTEVSWPCRNQSSSEKATNLGKKKSSWKAFFGVYGEKDSQSPKAVSAGQRT 119
DB 81 ICTDDEARKRTNAOC-----VCYKV-----WKHY--SAERKRGQP----- 114
OY 120 LEYDSSHQSSRCLSNVEQCLHEHADVDPKYISINRAEIVYSW-----PPQA 169
DB 115 -EVEGRHKDVFGR-----SQLRP-----SRLPREPELLFSGVYFESGLIPSLPLTL 159
OY 170 TQAG-----GFSKEIFYTEGLSFQLOGHVPVPAASSKKRDEEQ 207
DB 160 SQCGRNCILVFQNGEQFASPSMISEQALQSLDHLHYQRYAEERKLEPQ 208

RESULT 14
O9EQS9 PRELIMINARY; PRT; 1252 AA.
AC O9EQS9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE DDM36
GN NOPE OR DDM36.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6JA;
RA Aoyama T., Nakamura T., Nakayama T., Yamamoto H., Hosaka T.,
RA Aoyama T., Nagayama S., Oka M., Kiyono T., Sasaki M.S., Nakamura T.,
RA Toguchida J.;
RT "Up-regulation of a ras effector and down-regulation of a cell
RT adhesion molecule are associated with transformation of osteoblasts.";
RT Submitted (DCC-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AB052620; BAB19278.1; -.
HGSP: P56276; IRLK.
MSD: MGI:1858497; Nope.
DR InterPro: IPR003962; FNIII_repeat.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR003599; Ig_2.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003600; Ig_like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00041; fn3_5.
DR Pfam: PF00047; Ig_4.
DR PRINTS: PRO0014; FNTPEIIT.
DR SMART: SMO0060; FN3_5.
DR SMART: SMO0409; IG_4.
DR SMART: SMO0408; IGC2_4.
DR SMART: SMO0410; IG_like_2.
DR Immunoglobulin domain, Repeat.
SQ SEQUENCE 1252 AA; 134764 MW; BA52932939483AB73 CRC64;

```

Query Match	6.8%	Score 89;	DB 11;	Length 1252;
Best Local Similarity	23.7%	Pred. No. 22;		
Matches 46;	Conservative 26;	Mismatches 82;	Indels 40;	Gaps 8;
QY	8 DLEPRLDDDLNLT-IEKILAYIRRHVFSPTALFS-----PKLLTRRSLSQSGLCNC	61		
	: : : : : :	:		:

```
Db      483 DTTELQVADLEPNPDIETYYVAAYSQLGASRRSSPALHTLDDVPASAPQTLLS----- 535  
Oy      62 SANESWTEVSVM-PCRNSSOSEKAINLGKKRSSWKAFGEYKEEDSOSTPAKVSAOGRTL 120  
         | : | : | : | : | : | : | : |  
Db     536 SPNPEDIIRAVAWLPSPSLISNGQVL-----KYKIEXG-LCKEDQGVSTEVCNENDTLT 587  
Oy    121 EYDSSH-----QQMSRCLSNVEQCILEHAVDPKIYSIANRYALVIYS 163  
         :                || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db     588 NSLDPNKNKYRVARISAGTGAGYGVSQMWHRTPGVNQHSDHVFPEAIKLVRAKMESLYVS 647  
Oy    164 W-PPOATQGAGEF 176  
         ||||| |||||  
Db     648 WOPEHPPTQIGSYK 661
```

RESULT 15

O9JULI PRELIMINARY; PRT: 1252 AA.

ID O9JULI
AC Q9JULI;
DT 01-OCT-2000 (TREMBlrel_15, Created)
DT 01-OCT-2000 (TREMBlrel_15, last sequence update)
DE 01-DEC-2001 (TREMBlrel_19, last annotation update)
NEIGHBOR OF PUNC EII PROTEIN.
GN NOPE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Mus.
OX NCBI_TaxID=10090;
XP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB;
RX MEDLINE=20175427; PubMed=10708514;
RA Salbaum J.M., Kappen C.;
RT "Cloning and expression of Nope, a new mouse gene of the
immunoglobulin superfamily related to guidance receptors.";
RL Genomics 64:15-23(2000).
DR EMBL; AF176694; AAF65930.1; -.
DR HSSP; P56276; ITLK.
DR MGD; MG1:1858497; Nope.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003968; Ig_c2.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00041; Fn3_5.
DR Pfam; PF00047; Ig_4.
DR PRINTS; PR00014; FNTYPEIII.
DR SMART; SMO0060; FN3; 5.
DR SMART; SMO040B; IGc2; 3.
DR SMART; SMO0410; IG_Like; 1.
KW Immunoglobulin domain; Repeat.
SQ SEQUENCE 1252 AA; 134759 MW; 1194877327B7B76B2 CRC64;

Query Match	6.8%	Score 89	DB 11	Length 1252
Best Local Similarity	23.7%	Pred. No. 22		
Matches	46	Conservative	26	Mismatches 82; Indels 40; Gaps 8

QY	8	DLEELPLDDDDLTNTIEFKILAYTTRHHYFKSTPALFS----	PKLLRTSLRSISORGLGNC	61
		: : : :		
Db	483	DTTEQLQVRDLREPTDYEFEVYVAYSQLGASRTSSPALVHTLDDVPSAARQLTUS-----		535
QY	62	SANESWTEVSW-PCNNSOSSSEKAILGKKKSSWKMAFFGVGEKEDSDSQSTPAKYSAGQRTL		120
		: : : : : : :		
Db	536	SPNPEDIRAVMLPELPSSLSNGOVL-----	KYKLEIYG-LGKEDGVFSTVEPQGNELQTLT	587
QY	121	EYDSDHS-----	QOMSKCLSNVQCLSEHEAVDEKVISIANRVAEIIYS	163
		: : : : : :		
Db	588	NSLDQNKYRVRIASGTAGCYGVPSQMMQHRRITGVHNDQSHVFAPAEELKIVRAKMSLIYVS		647
QY	164	W-PPRQATQAGGFK	176	
Db	648	WQRPHPPTQISGK	661	

Thu Jun 20 13:15:11 2002

us-09-771-961-4.rspt

Page 8

Search completed: June 19, 2002, 16:39:46
Job time: 399 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2002, 16:40:06 ; Search time 15.65 Seconds

(without alignments)
623.471 Million cell updates/sec

Title: US-09-771-961-4

Perfect score: 1316

Sequence: 1 MCTSGCDLEIPLDDDLN.....IPLVDTSIQGFPPDGLMACI 252

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	101.5	7.7	530	1 TDT_MOUSE	P09638 mus musculus
2	91.5	7.0	2319	1 ARA6_HUMAN	Q13023 homo sapien
3	90	6.8	433	1 KAR1_YEAST	P11927 saccharomyc
4	89.5	6.8	696	1 UVRG_DROME	Q9VX07 drosophila
5	88.5	6.7	1131	1 AC15_MOUSE	P35601 mus musculus
6	88	6.7	486	1 TECL1_YEAST	P18412 saccharomyc
7	87	6.6	520	1 TDT_BOVIN	P06526 bos taurus
8	86	6.5	639	1 GCAL_HUMAN	Q9UJ35 homo sapien
9	85.5	6.5	250	1 YC97_HUMAN	Q9Y345 homo sapien
10	85.5	6.5	464	1 CAP_DICDI	P54654 dictyostel
11	85	6.5	702	1 DY12_ANTCR	Q16959 antioxiidari
12	85	6.5	1468	1 DPOA_YEAST	P13382 saccharomyc
13	84.5	6.4	2875	1 RRPL_TSWY1	P28976 tomato spot
14	83	6.3	1264	1 STV2_HUMAN	P26640 homo sapien
15	82.5	6.3	1468	1 G1ND_ECOLI	P49791 rattus norv
16	81	6.2	890	1 GLND_YEAST	P27249 escherichia
17	80.5	6.1	1653	1 CLH1_YEAST	P22137 saccharomyc
18	80.5	6.1	1722	1 RBH2_HUMAN	P29375 homo sapien
19	80	6.1	177	1 R66_MITC1	P33099 micrococcus
20	80	6.1	690	1 STGB_BUCAL	P57335 buchnera ap
21	79.5	6.0	1242	1 MSH6_YEAST	Q03834 saccharomyc
22	79.5	6.0	2314	1 AKAG_RAT	Q09W07 rattus norv
23	79	6.0	445	1 FIDL_VIRPA	Q03475 vibrrio para
24	79	6.0	723	1 GGA3_HUMAN	Q9N552 homo sapien
25	79	6.0	900	1 SYA_MYCG	P47534 mycoplasma
26	78.5	6.0	184	1 CAS3_MOUSE	Q02862 mus musculus
27	78.5	6.0	822	1 YF1_YEAST	P47046 saccharomyc
28	78.5	6.0	1123	1 DC11_DROME	P18169 drosophila
29	78.5	6.0	1589	1 DC13_DROME	P18171 drosophila
30	78	5.9	758	1 YMS8_YEAST	Q03825 saccharomyc
31	78	5.9	1147	1 AC15_HUMAN	P35251 homo sapien
32	77.5	5.9	1777	1 BARI1_HUMAN	Q99128 homo sapien
33	77.5	5.9	1263	1 STV2_MOUSE	Q92199 mus musculus

34	77.5	5.9	1428	1 YAB4_SCHPO	Q09773 schizosacch
35	77.5	5.9	1636	1 BUD3_YEAST	P25558 saccharomyc
36	77	5.9	623	1 HPC2_YEAST	Q01448 saccharomyc
37	77	5.9	630	1 ACHA_RAT	P09483 rattus norv
38	77	5.9	906	1 PRO1_SCHPO	Q12381 schizosacch
39	77	5.9	1805	1 NEST_RAT	P21263 schizosacch
40	76.5	5.8	502	1 CP12_HUMAN	P51589 homo sapien
41	76.5	5.8	518	1 TDT_MONDO	Q02789 monodelphis
42	76	5.8	486	1 YN8U_YEAST	P53742 saccharomyc
43	76	5.8	572	1 GAG_IPHA	P04023 hamster int
44	76	5.8	597	1 SYK_AQUAE	Q67258 aquilex int
45	76	5.8	950	1 DC12_DROME	P18170 drosophila

ALIGNMENTS

RESULT 1

ID	TDT_MOUSE	STANDARD	PRT	530 AA.
AC	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-2002 (Rel. 41, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	DNA nucleotidylexotransferase (EC 2.7.7.31) (Terminal addition enzyme)			
DE	(Terminal deoxynucleotidyltransferase) (TDT) (Terminal transferase).			
GN	DNTT OR TDT.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=86286588; PubMed=3755527;			
RA	Kohwai O., Yokota T., Kageyama T., Hirose T., Yoshida S., Arai K.;			
RT	"Isolation and characterization of bovine and mouse terminal			
RT	deoxynucleotidyltransferase cDNAs expressible in mammalian cells."			
RL	Nucleic Acids Res. 14:5777-5792(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BAIB/C; TISSUE=Thymus;			
RX	MEDLINE=93219079; PubMed=8464703;			
RA	Doyen N., Fanton D'Andon M., Bentolila L.A., Nguyen T.Q., Rougeon F.;			
RT	"Differential splicing in mouse thymus generates two forms of			
RT	terminal deoxynucleotidyl transferase."			
RL	Nucleic Acids Res. 21:1187-1191(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND CHARACTERIZATION.			
RC	STRAIN=C57BL/6; TISSUE=Thymus;			
RX	Pubmed=11136823;			
RA	Benedict C.L., Gillfillan S., Kearney J.F.;			
RT	"The long isoform of terminal deoxynucleotidyl transferase (Tdtl)			
RT	enters the nucleus and, rather than catalyzing N addition, modulates			
RT	the catalytic activity of the short isoform."			
RL	J. Exp. Med. 193:89-99(2001).			
RN	[4]			
RP	PRELIMINARY CHARACTERIZATION OF ALTERNATIVE FORMS.			
RX	MEDLINE=96016194; PubMed=7556063;			
RA	Bentolila L.A., Fanton D'Andon M., Nguyen T.Q., Martinez O.;			
RT	"The two isoforms of mouse terminal deoxynucleotidyl transferase			
RT	differ in both the ability to add N regions and subcellular			
RT	localization."			
RL	EMBO J. 14:4221-4229(1995).			
CC	-I- FUNCTION: TEMPLATE-INDEPENDENT DNA POLYMERASE WHICH CATALYZES THE			
CC	RANDOM ADDITION OF DEOXYNUCLEOSIDE 5'-TRIPHOSPHATE TO THE 3' END OF			
CC	A DNA INITIATOR. ONE OF THE IN-VIVO FUNCTION OF THIS ENZYME IS THE			
CC	ADDITION OF NUCLEOTIDES AT THE JUNCTION (N REGION) OF REARRANGED			
CC	IG HEAVY CHAIN AND T CELL RECEPTOR GENE SEGMENTS DURING THE			
CC	MATURATION OF B AND T CELLS. The Tdtl isoform seems to serve to			
CC	modulate the function of Tdts.			
CC	-I- CATALYTIC ACTIVITY: N = N diphosphate + (deoxynucleotide)(M+N).			
CC	{deoxynucleotide}(M) = N diphosphate + (deoxynucleotide)(M+N).			

CC -1- CORFACTOR: REQUIRES MAGNESIUM.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, TDF-L(arge)/TDF-L (shown here)
 CC and TDF-S(mall)/TDF-S are produced by alternative splicing. The
 CC TDF-S form is the major form.
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-X FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BRCT DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X04123; CAA27735.1; -
 CC EMBL: X68670; CAA48634.1; -
 CC EMBL: AF316014; AAK07884.1; -
 CC EMBL: AF316015; AAK07885.1; -
 CC PIR: B23595; B23595.
 CC HSSP: P06766; 1BPB.
 CC MGD: MG1:98659; Dntf.
 CC InterPro: IPR001357; BRCT.
 CC InterPro: IPR002054; DNA-POLX.
 CC Pfam: PF00533; BRCT; 1.
 CC Pfam: PF00966; DNA-PolymeraseX; 1.
 CC PRINTS: PR00869; DNAPOLX.
 CC SMART: SM00292; BRCT; 1.
 CC SMART: SM00483; POLXC; 1.
 CC PROSITE: PSS0172; BRCT; 1.
 CC PROSITE: PSS00522; DNA-POLYMERASE_X; 1.
 CC TRANSFAC: PS00522; DNA-POLYMERASE_X; 1.
 CC TRANSFAC: Nucleotidyltransferase; Terminal addition; Magnesium;
 CC Nuclear protein; Alternative splicing.
 CC
 CC FT DOMAIN 27 118
 CC FT VARSPIC 483 502 MISSING (IN ISOFORM TDF-S).
 CC FT CONFLICT 26 26 T -> M (IN REF. 2).
 CC FT CONFLICT 99 99 L -> F (IN REF. 2).
 CC FT CONFLICT 193 193 R -> G (IN REF. 1).
 CC FT CONFLICT 287 287 Q -> K (IN REF. 1).
 CC FT CONFLICT 309 309 E -> O (IN REF. 1).
 CC FT CONFLICT 367 367 D -> H (IN REF. 1).
 CC FT CONFLICT 441 444 DRRA -> ECAC (IN REF. 1).
 CC FT CONFLICT 443 445 RAF -> AS (IN REF. 2).
 CC FT CONFLICT 530 AA; 60331 MW; EBB109DCP39C8107 CRC64;
 CC SEQUENCE

Query Match 7.7%; Score 101.5; DB 1; Length 530;
 Best Local Similarity 23.5%; Pred. No. 0.28;
 Matches 54; Conservative 41; Mismatches 92; Indels 43; Gaps 11;

8 DLEIPDDDDLTITFEKILAYTRHNFKSTPALFSPKLTFRSHSQGLGNCNSANESW 67
 170 DADIDLAENDELRENESGCLAFMASSVLSKSLPRPI-----TSMKDTGI -PCLDQVK 222
 68 TEVSWPCRNSSSE-KAINIGKKKSSWKAFF--EGVEKEDSQSTPAKVSAGQRTL-EX 122
 223 SIEGIIIEDESSCAKAVLNDERKSKFKLTISVFGVGLK-----TAEKMFRRMGFRLSKI 277
 123 QDSHSQMSNC-----LSNVEOCLHEAVDPKYISIANRAVEIYSW-PPQAQAQAG 174
 278 QSKRSLEFTQAKAGFLYEYELVSCVNR-----PEAEVASMLKVAVTELPDALVTMTG 333
 175 FKSKEIFVTEGLSFOLOGH---VPVASSKRDDEEOILAKIVELLKYSG 220
 334 FRNG-----KMTGHVDPLITSPEAIEDEEOQLHKYTDWKQOG 373

RESULT 2
 AKA6_HUMAN STANDARD; PRT; 2319 AA.
 ID AC 013023; 015028;
 DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE A-kinase anchor protein 6 (Protein kinase A anchoring protein 6)
 DE (AKAP6) (A-kinase anchor protein 100 kDa) (AKAP 100) (MAKAP).
 GN AKAP6 OR AKAP100 OR KIAA0311.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 CC NCBI_Taxid=9606;
 CC [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE-99343692; PubMed-10413680;
 RA Kapiloff M.S., Shillace R.V., Westphal A.M., Scott J.D.;
 RT "MAKAP: an A-kinase anchoring protein targeted to the nuclear membrane
 RT of differentiated myocytes.";
 RL J. Cell Sci. 112:2725-2736(1999).
 RN [2]
 RP SEQUENCE OF 934-2319 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE-97349984; PubMed-9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [3]
 RP SEQUENCE OF 1666-2319 FROM N.A.
 RC TISSUE=Hippocampus;
 RX MEDLINE-95238446; PubMed-7721854;
 RA McCartney S., Little B.M., Langeberg L.K., Scott J.D.;
 RT "Cloning and characterization of A-kinase anchor protein 100
 RT (AKAP100). A protein that targets A-kinase to the sarcoplasmic
 RT reticulum.";
 RL J. Biol. Chem. 270:9327-9333(1995).
 CC -1- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE A
 CC AND ANCHORS/TARGETS THEM TO THE NUCLEAR MEMBRANE OR SARCOPLASMIC
 CC RETICULUM. MAY ACT AS AN ADAPTER FOR ASSEMBLING MULTIPROTEIN
 CC COMPLEXES.
 CC -1- SUBUNIT: INTERACTS WITH RII SUBUNIT OF PKA, PHOSPHATASE 2B
 CC (CALCINEURIN) AND AKAP79.
 CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM AND NUCLEAR MEMBRANE
 CC IN HEART MUSCLE. PARTICIPATION OF MULTIPLE TARGETING SIGNALS ALLOW
 CC CORRECT INTRACELLULAR TARGETING. THESE MAY BE REPEATED MOTIFS RICH
 CC IN BASIC AND HYDROPHOBIC AMINO ACIDS, PALMITOYLATED/MYRISTOYLATED
 CC MOTIFS OR ALTERNATIVELY SPLICED TARGETING SEQUENCES.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN CARDIAC AND SKELETAL
 CC MUSCLE, FOLLOWED BY BRAIN.
 CC -1- DOMAIN: RII-ALPHA BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC
 CC HELIX, COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
 CC -1- SIMILARITY: CONTAINS 2 SPECTRIN REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U17195; AAA92354.2; -
 CC EMBL: AB002309; BAA20770.1; -
 CC MIM: 604691; -
 CC InterPro: IPR002017; Spectrin.
 CC Pfam: PF00435; Spectrin; 2.
 CC SMART: SM00150; SPEC; 2.
 CC Repeat.
 CC REPEAT 762 848 SPECTRIN 1.
 CC REPEAT 1036 1150 SPECTRIN 2.
 CC DOMAIN 1560 1701 SER-RICH.
 CC DOMAIN 2063 2076 PKA-RII SUBUNIT BINDING DOMAIN.
 CC CONFLICT 974 974 C -> W (IN REF. 2).

RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibeagwa C.,
RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervlov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh K.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000)
CC -1- SIMILARITY: TO MAMMALIAN UVRAG.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE03639; AAF53277.1; -
DR Flybase: FBgn0032499; CG6116.
KW Hypothetical protein; Coiled coil.
FT DOMAIN 249 282 COILED COIL (POTENTIAL).
FT DOMAIN 329 398 COILED COIL (POTENTIAL).
SQ SEQUENCE 696 AA; 80690 MW; 26FDA367CF8D0CD CRC64;

Query Match 6.84; Score 89.5; DB 1; Length 696;
Best Local Similarity 21.08; Pred. No. 4.3;
Matches 48; Conservative 29; Mismatches 77; Indels 75; Gaps 11;

QY 6 GCDLEIP-----LDDDDLTIEFKLAAAYTRHHVFKSTPALFSPKILRTSLRSRGIG 59
DB 28 GFNIESMPHKSLELDDPD-----DVLTYTL-HTDKASEFPTSEKLPQKHQOKMAE 80
QY 60 NCSANESWTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVYKEKDSQSTPARVSAQOQRT 119
DB 81 ICTDDDEAMRKTNAC-----VCVKV-----WKHY--SABRRDQDP----- 114
QY 120 LEYQDSHQSQMSKCLSNVEQCLEHADVPRKYISIANRAVEIYSW-----PPQA 169
DB 115 -EVEQRHKDVEGR-----SQLTP-----SRLPRELLFSGVYFSGILPISPLTL 159
QY 170 TQAG-----GFKSKEIYEFGSLFQLOGHVPAASSKSKDEEQ 207
DB 160 SQCGRNCIVQLNGEOPASPSMISEQALQSLDHLHYQKAYAEKKELEEPQ 208

RESULT 5
AC15_MOUSE STANDARD; PRT; 1131 AA.
AC P35601:
DR 01-JUN-1994 (rel. 29, Created)
DR 01-OCT-1996 (rel. 34, Last sequence update)
DR 16-OCT-2001 (rel. 40, Last annotation update)
DE Activator 1 140 kDa subunit (Replication factor C large subunit) (A1
DE 140 kDa subunit) (RF-C 140 kDa subunit) (Activator 1 large subunit)
DE (A1-P145) (Differentiation specific element binding protein)
DE (ISRE-binding protein).
GN FRC1 OR RECC1 OR IBF-1.

OS Mus musculus (Mouse).
CC Eukaryota; Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX MGI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA16/C;
RX MEDLINE=94089669; PubMed=826586;
RA Burdello P.D., Utani A., Pan Z., Yamada Y.,
RT "Cloning of the large subunit of activator 1 (replication factor C)
RT reveals homology with bacterial DNA ligases.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11543-11547(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94158835; PubMed=8114700;
RA Luckow B., Bunz F., Stillman B., Lichter P., Schuetz G.,
RT "Cloning, expression, and chromosomal localization of the
RT 140-kilodalton subunit of replication factor C from mice and
RT humans.";
RL Mol. Cell. Biol. 14:1626-1634(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS;
RX MEDLINE=95388065; PubMed=7659092;
RA McGhee Habener J.F.;
RT "Differentiation-specific element binding protein (DSEB) binds to a
RT defined element in the promoter of the angiotensinogen gene required
RT for the irreversible induction of gene expression during
RT differentiation of 3T3-L1 adipoblasts to adipocytes.";
RL Mol. Endocrinol. 9:487-501(1995).
RN [4]
RP SEQUENCE FROM N.A.
RA Haque S.J.;
RL Submitted (FEB-1994) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-565 FROM N.A.
RC STRAIN=LAFL;
RA Lossie A.C., Haugen B.H., Wood W.M., Camper S.A., Gordon D.F.;
RL Submitted (OCT-1994) to the EMBL/Genbank/DBJ databases.
RN [6]
RP SEQUENCE OF 354-528 FROM N.A.
RX MEDLINE=90229765; PubMed=1691767;
RA Haque S.J., Kumar A., Fischer T., Rutherford M.N., Williams B.R.;
RT "Evaluation of inter- and intramolecular primary structure homologies
RT of interferons by a Monte Carlo method.";
RL J. Interferon Res. 10:31-31(1990).
CC -1- FUNCTION: THE ELONGATION OF PRIMED DNA TEMPLATES BY DNA POLYMERASE
CC DELTA AND EPSILON ROUTINES THE ACTION OF THE ACCESSORY PROTEINS
CC PCNA AND ACTIVATOR 1. THE 140 SUBUNIT BINDS TO THE PRIMER-
CC TEMPLATE JUNCTION.
CC -1- SUBUNIT: HEMEROPENTAMER OF SUBUNITS OF 140/145, 40, 38, 37, AND
CC 36.5 KDA THAT FORMS A COMPLEX WITH PCNA IN THE PRESENCE OF ATP.
CC -1- SIMILARITY: BELONGS TO THE ACTIVATOR 1 140 KDA SUBUNIT FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BRC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U01222; AAA21643.1; -
DR EMBL: X72711; CAA51260.1; -
DR EMBL: U36441; AAA79698.1; -
DR EMBL: U07157; AAS52140.1; -
DR EMBL: U15037; AAB60452.1; -
DR MGI: 97891; Reccl.
DR InterPro: IPR003959; AAA_subfam.
DR InterPro: IPR001357; BRC.

DR InterPro: IPR000862; RRC.
 DR Pfam: PF00004; AAA; 1.
 DR Pfam: PF00533; BRC1; 1.
 DR SMART: SM00292; BRC1; 1.
 DR PROSITE: PS00172; BRC1; 1.
 DR DNA replication; ATP-binding; Transcription regulation; DNA-binding;
 KW Activator; Nuclear protein; Zinc-finger.
 FT DOMAIN 399 477
 FT NP_BIND 635 642
 FT ZN_FING 734 751
 FT DOMAIN 1104 1108
 FT DOMAIN 354 528
 FT CONFLICT 66 66
 FT CONFLICT 187 187
 FT CONFLICT 254 254
 FT CONFLICT 559 559
 FT CONFLICT 614 614
 FT CONFLICT 945 945
 FT CONFLICT 1071 1071
 FT CONFLICT 1104 1104
 SO SEQUENCE 1131 AA; 125984 MW; A6FAF970A7FE9E94 CRC64;

Query Match 6.7%; Score 88.5; DB 1; Length 1131;
 Best Local Similarity 23.7%; Pred. No. 9.7;
 Matches 58; Conservative 34; Mismatches 92; Indels 61; Gaps 11;

QY 48 LRRSLSORGL-----GNCNSNESWTEVSW-----PCNSQSSEKAINIGKR 89
 DB 462 LGRIKIDEDGLDIRTPWPKRSKYEAAEMKREKSKLETPQKNDGKR-ISPARK 520
 QY 90 KSSWKAFFGVKEKSDSTPAKVSAGORTLEYODSHSOQMSRCLSNVEQCL-----EHEA 145
 DB 521 EESKCKKCLTLKNSPKAKVKEASTCPRGIDYKETHGR-----SSNKECLLWVDKYP 576
 QY 146 VDPKVT-----SIANVAEIVYSW-----PPQATQAGGKSEIFVTEGLSFQ--- 189
 DB 577 ASLKNIIIGQOGDQSCANLLRLWLRNMHKSPEEKRAAKFKGLASKD---DGSFKAAL 632
 QY 190 LGHVPAVSSSKKDEEGLIAKIVELLKYSQDLERKDT-----AFIPILVDSISQ 242
 DB 633 LSGPPGVGKTT-----TASLVCOELGYIVELNASTDRSKNSLKAVAESINNTSIKG 685
 QY 243 FPQDG 247
 DB 686 FYTSG 690
 RESULT 6
 TECL_YEAST STANDARD; PRT: 486 AA.
 ID TECL_YEAST
 AC P18412;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE TY transcription activator TECL.
 GN TECL OR ROCI OR YBR083W OR YBR0750.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RX MEDLINE=90287143; PubMed=2192259;
 RA Laloux I., Dubois E., Dewechin M., Jacobs E.;
 RT "TECL, a gene involved in the activation of TYI and TYI-mediated gene
 expression in Saccharomyces cerevisiae: cloning and molecular
 analysis.";
 RL Mol. Cell. Biol. 10:3541-3550(1990).
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;

RA Andre B., Cziepluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
 RA Vissers S.;
 RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP TEA DOMAIN.
 RX MEDLINE=91300541; PubMed=2070413;
 RA Buerklin T.R.;
 RT "The TEA domain: a novel, highly conserved DNA-binding motif";
 RL Cell 66:11-12(1991).
 CC -1- FUNCTION: TECL IS INVOLVED IN THE ACTIVATION OF TYI AND TYI-
 CC MEDIATED GENE EXPRESSION. IT IS NOT INVOLVED IN MATING OR
 CC SPOULATION PROCESSES.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: CONTAINS 1 'TEA' DNA-BINDING DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: M32797; AAA35141.1; -
 DR EMBL: Z35952; CAA85028.1; -
 DR PIR: A35667; A35667.
 DR TRNSPAC: T01084; -
 DR SGD: S0000287; TECL.
 DR InterPro: IPR000818; TEA.
 DR Pfam: PF01285; TEA; 1.
 DR PRINTS: PR00065; TEADOMAIN.
 DR SMART: SM00426; TEA; 1.
 DR PROSITE: PS00554; TEA.DOMAIN; 1.
 KW Transcription regulation; Trans-acting factor; Activator;
 KW DNA-binding; Nuclear protein. TEA.DOMAIN.
 FT DNA_BIND 127 192
 SO SEQUENCE 486 AA; 55157 MW; F247016D3E75C454 CRC64;

Query Match 6.7%; Score 88; DB 1; Length 486;
 Best Local Similarity 22.6%; Pred. No. 3.6;
 Matches 58; Conservative 36; Mismatches 95; Indels 68; Gaps 12;

QY 3 STSGCDLEIPDDDDINTTEFKILAYTRHHYK-----STPALFSPK 46
 DB 48 STACVKNKA---EDNINILD-----THPGEIVNTGIGAKSDLKSPAKFTFDK 96
 QY 47 LIRTR-----SLSGRGICGNSANSWTEVSWPCRNSSQSEK-----AINIGKKSSMK- 94
 DB 97 QRKNEVPNITSVSNYPFGQSEETSSITE-SWTIGCDKSEKVEAFLEALRLIMKNGTKKI 155
 QY 95 ----AFEGVKEKSDSTPAKVSAGORTLEYODSHSOQMSRCLSN-VEQCLEHRAVPK 149
 DB 156 KIRNANFG---RNELISLYIKTKHKTNEPRIRKQJSSHIQWVKTKIQKIDSLTSSKEE 212
 QY 150 VISIANRVA-----ETVSWPPQATQAGGKSEIFV---TEGLSFQ 191
 DB 213 LHLHLEHGAEDQTENSNLFDFEELIDSL--PSVSDSGSLTPKNIIVYNNSSGSLSVSK 270
 QY 192 GHVPAVSSSKKDEEQI 208
 DB 271 LTPITASNEKKIENFI 287
 RESULT 7
 TDT_BOVIN STANDARD; PRT: 520 AA.
 ID TDT_BOVIN
 AC P06526;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA nucleotidylexotransferase (EC 2.7.7.31) (Terminal addition enzyme)
 DE (Terminal deoxynucleotidyltransferase) (TDT) (Terminal transferase).
 DE

DNT OR TDT.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euarchonta; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus.
RX MEDLINE=86286588; PubMed=3755527;
RA Koital O., Yokota T., Kagayama T., Hirose T., Yoshida S., Arai K.;
RT "Isolation and characterization of bovine and mouse terminal
deoxynucleotidyltransferase cDNAs expressible in mammalian cells.";
RL Nucleic Acids Res. 14:5777-5792(1986).
RN [2]
RP SEQUENCE OF 1-13 FROM N.A.
RX MEDLINE=87213162; PubMed=3579900;
RA Koital O., Kaneda T., Morishita R.;
RT "Analysis of human terminal deoxynucleotidy] transferase cDNA
expressible in mammalian cells".
RL Biochem. Biophys. Res. Commun. 144:185-190(1987).
RN [3]
RP PARTIAL SEQUENCE OF 221-249.
RX MEDLINE=88153669; PubMed=3346221;
RA Pandey V., Modak M.J.;
RT "Biochemistry of terminal deoxynucleotidyltransferase. Affinity
labeling and identification of the deoxyribose triphosphate
binding domain of terminal deoxynucleotidyltransferase";
RJ Biol. Chem. 263:3744-3751(1988).
CC -1- FUNCTION: TEMPLATE-INDEPENDENT DNA POLYMERASE WHICH CATALYZES THE
RANDOM ADDITION OF DEOXYNUCLEOSIDE 5'-TRIPHOSPHATE TO THE 3' END OF
A DNA INITIATOR. ONE OF THE IN-VIVO FUNCTION OF THIS ENZYME IS THE
ADDITION OF NUCLEOTIDES AT THE JUNCTION (N REGION) OF REARRANGED
IG HEAVY CHAIN AND T CELL RECEPTOR GENE SEGMENTS DURING THE
MUTATION OF B AND T CELLS.
CC NATURATION OF B AND T CELLS.
CC -1- CATALYTIC ACTIVITY: N deoxyribose triphosphate +
[deoxynucleotide](M) = N diphosphate + {deoxynucleotide}(M+N).
CC -1- COFACTOR: REQUIRES MAGNESIUM.
CC -1- SIMILARITY: BELONGS NO DNA POLYMERASE TYPE-X FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BRCT DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isdb.ch/announce/
or send an email to license@isdb.sib.ch).

EMBL: X04122; CAA27734.1; --
EMBL: M26146; AAA87354.1; --
PIR: A23595; A23595
PIR: A23948; A23948
InterPro: IPR001357; BRCT.
DR InterPro: IPR002054; DNA_POLX.
DR Pfam: PF00533; BRCT_1.
DR Pfam: PF00966; DNA_polymerase; 1.
DR PRINTS: PR00869; DNAPOLX.
SMART: SM00292; BRCT; 1.
SMART: SM00483; POLXc; 1.
DR SMART: PS50172; BRCT; 1.
DR PROSITE: PS00522; DNA_PLYMERASE_X; 1.
KW Transferase; Nucleotidyltransferase; Terminal addition; Magnesium.
FT DOMAIN 38 135 BRCT.
SEQUENCE 520 AA; 59670 MW; 258FE25B4E77A390 CRC64;

Db	234	C1EEI	---	IEDGESESEVNAV	UNDER	QSPKLT	SVF	-----	GVG	LTK	ISEK	276
OY	67	WTEVSWPCNSOSSSEKAINLGK	KKSSKMAFEGV	EKEEDSOSTPAK	VSAGORPLEY	ODSH	126					
Db	277	WFMGRSLSKIMSDTKLFTK	QKQ	---	KAGF	-----	LYEDLV	311				
OY	127	SQOWMSFCLSNVQCLHEHVA	DPVVISIANVAVELYSW	-PPQATQAGF	-KSKEIFVTE	184						
Db	312	S	---	CVTRAEAE	---	EAV	-----	GVLYKRAVMAFLDPALFTYMTGFRFRGK	---	351		
OY	185	GLSFOLGCH	---	VPVYSSSKDEEEOILAK	IVELTKYSG	-----	DOLEKRDYAE	-IP	IPL	235		
Db	352	-----	GHDVDFLITSPGSADEED	EDLPVINVLMWEKGLLY	VDYVESFTEFKLP	SRQ	404					
OY	236	VDT	238									
Db	405	VDT	407									

RESULT	8
GGAL_HUMAN	STANDARD; PRT; 639 AA.
ID	GGAL_HUMAN
AC	Q9UJY5; Q9UGW0; Q9UGW1; Q9UG00; Q9BW94;
DT	16-OCT-2001 (Rel. 40; Created)
DT	16-OCT-2001 (Rel. 40; Last sequence update)
DT	16-OCT-2001 (Rel. 40; Last annotation update)
DE	ADP-ribosylation factor binding protein GGAL (Golgi-1-localized, gamma ear-containing, ARF-binding protein 1) (gamma-adaptin related protein 1).
DE	GGAL.
GN	Homo sapiens (Human).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX	NCBI_Taxid=9606;
LN	[1]
RP	SEQUENCE FROM N.A. (ISOFORM 1).
RX	MEDLINE=20214818; PubMed=107449927;
RA	Boman A.L., Zhang C.-J., Zhu X., Kahn R.A.;
RT	"A family of ADP-ribosylation factor effectors that can alter transport through the trans-Golgi.";
RL	Mol. Biol. Cell 11:1241-1255(2000).
LN	[2]
RP	SEQUENCE FROM N.A. (ISOFORM 1).
RC	TISSUE=Heart;
RC	MEDLINE=20211638; PubMed=10747089;
RA	Dell'Angelica E.C., Puertollano R., Mullins C., Aguilar R.C.,
RA	Hirst J., Lui W.-Y., Bright N.A., Totty N., Seaman M.N.J.,
RA	Robinson M.S.;
RT	"A family of proteins with gamma-adaptin and VHS domains that facilitate trafficking between the trans-Golgi network and the vacuole/lysosome.";
RL	J. Cell Biol. 149:67-80(2000).
LN	[3]
RP	SEQUENCE FROM N.A. (ISOFORM 1).
RX	MEDLINE=20211637; PubMed=10747088;
RA	Hirst J., Lui W.-Y., Bright N.A., Totty N., Seaman M.N.J.,
RA	Robinson M.S.;
RP	PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA	Matthews L.;
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
LN	[5]
RP	SEQUENCE OF 31-639 FROM N.A. (ISOFORM 1).
RC	TISSUE=Testis;
RC	Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
LN	[6]
RP	SEQUENCE OF 272-369 FROM N.A.
RC	TISSUE=Neuroblastoma;
RA	Strausberg R.;
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

CC -I- FUNCTION: MAY HAVE A REGULATORY BIFUNCTIONAL ROLE. BINDS G-ACTIN
 CC AND PIP2. INVOLVED IN MICROFILAMENT REORGANIZATION NEAR THE PLASMA
 CC MEMBRANE IN A PIP2-REGULATED MANNER.
 CC -I- SUBCELLULAR LOCATION: CELL MEMBRANE.
 CC -I- DOMAIN: THE C-TERMINUS IS RESPONSIBLE FOR SEQUESTERING G-ACTIN.
 CC THE N-TERMINUS IS REQUIRED FOR THE PIP2 MODULATION OF CAP
 CC FUNCTION.
 CC -I- SIMILARITY: BELONGS TO THE CAP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U43027; AAB09713.1; -.
 CC DR DictyDb; DD00012; -.
 CC DR DictyPro; IPRO01837; CAP.
 CC DR Pfam; PF01213; CAP. 1.
 CC DR PROSITE; PS01088; CAP_1; 1.
 CC DR PROSITE; PS01089; CAP_2; 1.
 CC KW Membrane; Actin-binding.
 CC FT DOMAIN 33 42 POLY-SER.
 CC FT DOMAIN 217 253 PRO-RICH.
 CC FT SEQUENCE 464 AA; 49641 MW; 7ABDD05D8A0148B4 CRC64;

Query Match 6.5%; Score 85.5; DB 1; Length 464;
 Best Local Similarity 23.8%; Pred. No. 5.6;
 Matches 57; Conservative 31; Mismatches 99; Indels 53; Gaps 11;

38 STPAL-FSPKLTFRSLRSORG-----LGNCSSANSEWTEVSPCRNSQSSSEKAINIGKKK 90
 DB 232 SARAPAPAPAVSSITPVESSKKGPGIGAVGELSGDGVITSGKKVINDMSK---NFTDKS 288
 QY 91 SSMKAFEGVEEKEDSOSTPAKVSAGOR-TLEYODSH-----SOOWSRCLSNVEOCLEHE 144
 DB 289 SVAKADTVAKVADPAPRPVAFALQGNKWSLEYGVNKEIVIAEPDSRQTVYIFQCYN-- 346
 QY 145 AVDPKVIYSINRV-----AEIVSWPPPOATQAGGFSKELFTVEGLSFQLOGHV 194
 DB 347 ----SLVQIKGVNAITLDGCKKTSIVFE-----NAISSCEVNVNNGVEIQVTGRV 393
 QY 195 PVASSSK-----KD--EEEOILAKIVELTKYSGDLERKDTAFPIRP-LVDSISIOG 242
 DB 394 PSIAIDKTSQCIYLSKDSLETFEIVSSKSEMANVLIPGATENDDLVELAIPEQYKTSVKG 453
 ID 11
 ID 11 ANTCR STANDARD; PRT; 702 AA.
 AC 016959;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dynein intermediate chain 2, ciliary.
 DE Anthracidaris crassispina (Sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinozoa; Euechinozoa; Echinodermata; Echinozoa; Echinometridae;
 OC Anthracidaris.
 OC NCBI_TaxID=7629;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Egg;
 RA MEDLINE=96059480; PubMed=7579688;
 RA Ogawa K., Kamiya R., Wilkerson C.G., Witman G.B.;
 RT "Interspecies conservation of outer arm dynein intermediate chain
 RT sequences defines two intermediate chain subclasses."
 RL Mol. Biol. Cell 6:685-696(1995)
 RN -I- FUNCTION: MICROTUBULE-BINDING PROTEIN THAT MAY BE INVOLVED IN
 CC DYNEIN OUTER ARM ASSEMBLY ON THE AXONEME (BY SIMILARITY).

CC -I- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS (ALPHA AND BETA),
 CC THREE INTERMEDIATE CHAINS AND SEVERAL LIGHT CHAINS.
 CC -I- SIMILARITY: BELONGS TO THE DYNEIN INTERMEDIATE CHAIN FAMILY.
 CC -I- SIMILARITY: CONTAINS 6 WD REPEATS (TIR-ASP DOMAINS).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D38538; BAA07539.1; -.
 CC DR DictyPro; IPRO01680; WD40.
 CC DR Pfam; PF00400; WD40. 5.
 CC DR SMART; SM00320; WD40. 4.
 CC DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
 CC DR PROSITE; PS50082; WD_REPEATS_2; 1.
 CC DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 CC KW Motor protein; Microtubules; Dynein; Flagella; Repeat; WD repeat.
 CC FT REPEAT 380 420 WD 1.
 CC FT REPEAT 429 472 WD 2.
 CC FT REPEAT 490 533 WD 3.
 CC FT REPEAT 537 577 WD 4.
 CC FT REPEAT 580 620 WD 5.
 CC FT REPEAT 628 667 WD 6.
 CC FT SEQUENCE 702 AA; 79138 MW; 96931CFDAERBA8A CRC64;

Query Match 6.5%; Score 85; DB 1; Length 702;
 Best Local Similarity 25.6%; Pred. No. 11;
 Matches 33; Conservative 21; Mismatches 49; Indels 26; Gaps 4;

101 EKEDSOSTPAKVSAGORLTLEYQDSHQMSRCLSNV--EQCLEHEAVDPKVIYSINARVA 158
 DB 171 EKREDESTPPAPAPAKSDOKLTNOFNFSERASQYNNPFRGRQTE----- 216
 QY 159 EIVSWPPPOATQAGGFSKELF--VTEGLSFQLOGHVPAVSSSKKDEEOILAKIVEL 216
 DB 217 -----PPPRANFSSITANWEIYDAYMEDLERKEKKEKKAAPSKKDDKS--KKLTLAL 268
 QY 217 KYSGDLER 225
 DB 269 ETQSDMSR 277
 ID 12
 ID 12 DPOA_YEAST STANDARD; PRT; 1468 AA.
 AC P13382;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA polymerase alpha catalytic subunit (EC 2.7.7.7) (DNA polymerase
 DE 1).
 DE POL1 OR CDC17 OR YNL102W OR N2181.
 GN Saccharomyces cerevisiae (Baker's Yeast).
 OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC NCBI_TaxID=4932;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=88234507; PubMed=3287376;
 RA Pitzagalli A., Valsasani P., Plevani P., Lucchini G.;
 RT "DNA polymerase I gene of Saccharomyces cerevisiae: nucleotide
 RT sequence, mapping of a temperature-sensitive mutation, and protein
 RT homology with other DNA polymerases."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:3772-3776(1988).
 RN 12
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / FT1679;
 RX MEDLINE=96267765; PubMed=8701612;

RA Saiz J.E., Bultrao M.J., Soler A., del Rey F., Revuelta J.L.:
RT "The sequence of a 21.3 kb DNA fragment from the left arm of yeast
RT chromosome XIV reveals LEU4, MET4, POL1, RAS2, and six new open
RL reading frames.";
RL Yeast 12:403-409(1996).
RP [3]
RP SEQUENCE OF 1-26 FROM N.A.
RC STRAIN-S288C;
RX MEDLINE=93188702; PubMed=8446029;
RA Mountain H.A., Bystrom A.S., Korch C.;
RT "The general amino acid control regulates MET4, which encodes a
RT methionine-pathway-specific transcriptional activator of
RT Saccharomyces cerevisiae.";
RL Mol. Microbiol. 7:215-228(1993).
CC -I- FUNCTION: POLYMERASE ALPHA IN A COMPLEX WITH DNA PRIME IS A
CC REPLICATIVE POLYMERASE.
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: J03268; AAA3488.1; -;
DR EMBL: Z50161; CAA9052.1; -;
DR EMBL: Z71378; CAA95978.1; -;
DR EMBL: Z12126; CAA7811.1; -;
DR PIR: A50211; A50211.
DR SGD: S0005046; POL1.
DR InterPro: IPR002064; DNA_pol_B.
DR Pfam: PF00136; DNA_pol_B_1.
DR Pfam: PF03104; DNA_pol_B_exo; 1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBg; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
RW Transferrase: DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Nuclear protein.
DR DNA_BIND 1246 1381 POTENTIAL.
VARIANT 493 493 G->R (IN TEMPERATURE SENSITIVE MUTANT).
CONFLICT 759 760 MI->IV (IN REF. 1).
SEQUENCE 1468 AA; 166808 MW; 50c9e03d2db95b5ae CRC64;

RESULT	13			
RRPL_TSWV1				
ID	RRPL_TSWV1	STANDARD:	PRF:	2875 AA.
AC	P28976;			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	RNA-directed RNA polymerase (EC 2.7.7.48) (L protein).			
GN	L.			
OS	Tomato spotted wilt virus (strain Brazilian Br-01) (TSWV).			
OC	Viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus.			
OX	NCBI_TaxId=36413;			
RP	[1]			
RA	SEQUENCE FROM N.A.			
RX	MEDLINE=91374019; PubMed=1895058;			
RA	de Haan P., Kornelink R., de Oliveira Resende R., van Poelwijk F.,			
RA	Peters D., Goldbach R.;			
RT	"Tomato spotted wilt virus L RNA encodes a putative RNA polymerase.";			
RL	J. Gen. Virol. 72:2207-2216(1991).			
CC	-1: CATALYTIC ACTIVITY: N nucleoside triphosphate +			
CC	(RNA)(N).			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D10066; BAA00955.1; -.			
DR	PIR; J01335; RRVUTM.			
KW	Transferase; Nucleotidyltransferase; RNA-directed RNA polymerase.			
QO	SEQUENCE 2875 AA: 331498 MW: 574B3B55CF0706FB CR664;			

```

Query Match Similarity      6.5% ; Score 85 ; DB 1 ; Length 1468;
Best Local Similarity      25.9% ; Pred. No. 27;
Matches      51; Conservative 29; Mismatches 93; Indels 24; Gaps 8;

OY 65 ESMTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVGEKEDSOSTPAKVAQOGRTLEYQD 124
Db 74 EWMREVD---NSSSDDTGTLASKDSRRKK--NTRKEDHOITDMLRTQHSKSTLAAH 127
OY 125 SHSQQWRCLSNVEQGC---EHEAVDPKVISIANVAETVYWSPPPOATVAGGFKS--KE 179
Db 128 KKSQKKSIPIDNDDILGFEFSEGEVEKRNILLPSLIRELWNSPTAE-----FKSSIKR 181
OY 180 IFVTEGLSPQ--IQGHVPVASSSKKDEEEOQLAKIVEL---LKYSQDOLERRDTAFIP 233
Db 182 VNGNDESSHDAQISKYKIPDSSSTDKYLEIESPLKLQSRKLRVANDVODLLDV--ENS 240
OY 234 PLYDSTIQGPPDGLMA 250
Db 241 PLYATRKQNVLODPTLLA 257

```

	Query Match	6.4%	Score 84.5;	DB 1;	length 2875;
	Best Local Similarity	22.0%	Pred. No.71;		
	Matches 40;	Conservative 40;	Mismatches 73;	Indels 29;	Gaps 7;
Dd	QY	80	SEKAINLTKKKSSKNKAF-----GVEKEKEDSOSTPAKVSAOGRTLEYQDSHQMKSRCLS	135	
	:	: :	: :	: :	: :
Dd	Db	1760	NENNLN---KMSSTKDFSKILPNLNKKNEDLYKSTRNKLIIDEDAVE----EDELTKKLAS	1812	
	:	: :	: :	: :	: :
Dd	QY	136	NVEOCLHEAV--DPKVISIAN-----RVAEIYSWPPQATGAGFKSKEIFVTEGLS	187	
	:	: :	: :	: :	: :
Dd	Db	1813	SLEMESVADIMIKNPETLLIAPLDRODPLQLFLFAYTSPSRNQLSNSQSTEKLMDLVRLR	1872	
	:	: :	: :	: :	: :
QY	QY	188	FOLQGHPVPVASSSKKDEEQILAKIVELLKTSGD-----OLEKDTAF-IPDPLVD	237	
	:	: :	: :	: :	: :
Db	Db	1873	SKARTFVNISSTVMKYEEENMEKRILELPDLDSYCSFKTCVNLVIKDVFNSMLPIILD	1932	
	:	: :	: :	: :	: :
QY	QY	238	TS 239		
	:	: :	: :	: :	: :
Db	Db	1933	SA 1934		
	:	: :	: :	: :	: :
	RESULT 14				
	SYV2_HUMAN				
ID	SYV2_HUMAN	STANDARD:	PRT: 1264 AA.		
AC	P26640; Q9UOM2;				
DT	01-AUG-1992 (Rel. 23, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	01-MAR-2002 (Rel. 41, Last annotation update)				
DE	Valyl-tRNA synthetase 2 (EC 6.1.1.9) (Valine--tRNA ligase 2) (VALRS 2) (G7A).				
GN	VARS2 OR G7A.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	[1]				
	SEQUENCE FROM N.A.				

RX MEDLINE-91378943; PubMed-1898367;
 RA Hsieh H.-L., Campbell R.D.;
 RT "Evidence that gene Gta in the human major histocompatibility complex
 RT encodes valyl-tRNA synthetase.";
 RN Biochem. J. 278:809-816(1991).
 [2]
 RN ERRATUM.
 RA Hsieh S.-L., Campbell R.D.;
 RN Biochem. J. 281:879-879(1992).
 [3]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Madan A., Qin S., Shaffer T., James R., Ratcliffe A.,
 RA Abbasi N., Dichtoff R., Loretz C., Madan A., Dors M., Young J.,
 RA Lasry S., Hood L.;
 RT "Sequence of the human major histocompatibility complex class III
 RT region";
 RN Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RA Shilina S., Tamiya G., Oka A., Inoko H.;
 RT "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.";
 RN Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 [5]
 RP SEQUENCE OF 201-1263 FROM N.A.
 RA MEDLINE-93154582; PubMed-8428657;
 RA Vilalta A., Donovan D., Wood L., Vogeli G., Yang D.C.H.;
 RT "Cloning, sequencing and expression of a cDNA encoding mammalian
 RT valyl-tRNA synthetase.";
 RN Gene 123:181-186(1993).
 CC -1- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
 CC + L-valyl-tRNA(Val).
 CC -1- ENZYME REGULATION: CAN BE REGULATED BY PROTEIN KINASE C-DEPENDENT
 CC PHOSPHORYLATION.
 CC -1- SUBUNIT: FORMS HIGH-MOLECULAR-MASS AGGREGATES WITH ELONGATION
 CC FACTOR 1.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -1- SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION
 CC FACTOR 1-GAMMA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X59303; CAA41990.1; -;
 DR EMBL: AF134726; AAD21819.1; -;
 DR EMBL: AF000503; BAB63303.1; -;
 DR EMBL: M98326; AAA81332.1; -;
 DR PIR: S17675; S17675.
 DR HSSP: P96142; IGAX.
 DR MIM: 604137; -;
 DR InterPro: IPR004046; GST_C.
 DR InterPro: IPR004045; GST_N.
 DR InterPro: IPR002300; tRNA-synt_1a.
 DR InterPro: IPR001412; tRNA-synt_1i.
 DR InterPro: IPR002303; tRNA-synt_val.
 DR Pfam: PF00043; GST_C; 1.
 DR Pfam: PF02798; GST_N; 1.
 DR Pfam: PF00133; tRNA-synt_1; 1.
 DR PRINTS: PR00986; TRNASYNTHAL.
 DR PROSITE: PS00178; AA.TRNA.LIGASE.I; 1.
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Phosphorylation.
 KM DOMAIN 1 7200
 FT SITE 344 354 EF-1-GAMMA LIKE.
 FT SITE 862 864 "HIGH" REGION.
 FT BINDING 865 865 "KMSKS" REGION.
 FT CONFLICT 51 51 ATP (BY SIMILARITY).
 FT CONFLICT 331 331 P -> S (IN REF. 1).
 FT CONFLICT 331 331 A -> G (IN REF. 5).
 FT CONFLICT 590 590 V -> G (IN REF. 1).
 CC -----

FT CONFLICT 620 640 ALINVPPEGLGPRFEARRAV -> GPHCASAFPGPAQYL
 FT CONFLICT 792 792 RPGRKC (IN REF. 1).
 FT CONFLICT 1064 1064 S -> F (IN REF. 1).
 FT CONFLICT 1169 1169 M -> I (IN REF. 5).
 FT CONFLICT 1169 1169 MISSING (IN REF. 5).
 SQ SEQUENCE 1264 AA; 140475 MW; 95CCDBB3AB148AD CRC64;
 Query Match 6.3%; Score 83; DB 1; Length 1264;
 Best Local Similarity 22.1%; Pred. No. 33;
 Matches 49; Conservative 35; Mismatches 86; Indels 52; Gaps 11;
 QY 31 TRHHVKSPTALPESP-----KLRTRSLSGRGNGCNSANSEMTVEVSPCR 75
 DB TRHHVKSPTALPESP-----KLRTRSLSGRGNGCNSANSEMTVEVSPCR 75
 QY 76 NSQSEKAINLCKKRS--W-KAFEGVKEKDSQSPRAKVSAGQRLLEQDSHQMSR 132
 DB NSQSEKAINLCKKRS--W-KAFEGVKEKDSQSPRAKVSAGQRLLEQDSHQMSR 132
 QY 421 EKGDRIVHQLKRLSSLDMDACFTMDPKLSAATFAVRLH-EEGIIYRSTFLVWMS 479
 DB EKGDRIVHQLKRLSSLDMDACFTMDPKLSAATFAVRLH-EEGIIYRSTFLVWMS 479
 QY 133 CLSNVEQCLEHNAVDPKVISINRVAEIYVMPPEQATQAGFKSK-EIFVTEGLSPQL 191
 DB CLSNVEQCLEHNAVDPKVISINRVAEIYVMPPEQATQAGFKSK-EIFVTEGLSPQL 191
 DB 480 TLNSAISDIE--VDKKEIT-----GRTLSIVP-----GKEKVEGVLSFAIKVQ 523
 QY 192 GHVPYASSSKKDEBQILAKIVELKYSGD---QLERKDTAF 230
 DB GHVPYASSSKKDEBQILAKIVELKYSGD---QLERKDTAF 230
 DB 524 G-----SDSDEVVAVATRIETMTLGDVAVVHPKDRY 556
 G-----SDSDEVVAVATRIETMTLGDVAVVHPKDRY 556
 RESULT 15
 N153_RAT
 ID N153_RAT STANDARD; PRT; 1468 AA.
 AC P49791;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Nuclear pore complex protein Nup153 (Nucleoporin Nup153) (153 kDa
 DE nucleoporin).
 GN NUP153.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 602-613; 622-645 AND 971-993.
 RC STRAIN=BUFAFO; TISSUE=Liver;
 RX MEDLINE-93137325; PubMed-8422679;
 RA Sukegawa J., Blobel G.;
 RT "A nuclear pore complex protein that contains zinc finger motifs,
 RT binds DNA, and faces the nucleoplasm.";
 RN Cell 72:29-38(1993).
 CC -1- FUNCTION: POSSIBLE DNA-BINDING SUBUNIT OF THE NUCLEAR PORE
 CC COMPLEX (NPC). THE REPEAT-CONTAINING DOMAIN MAY BE INVOLVED IN
 CC ANCHORING COMPONENTS OF THE PORE COMPLEX TO THE PORE MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX. LOCATED TO THE
 CC TERMINAL RING STRUCTURE OF THE NUCLEOPLASMIC CAGE.
 CC -1- DOMAIN: CONTAINS X-F-X-F-G REPEATS.
 CC -1- PFM: PHOSPHORYLATED IN INTERPHASE. HYPERPHOSPHORYLATED DURING
 CC MITOSIS. MAY PLAY A ROLE IN THE REVERSIBLE DISASSEMBLY OF THE
 CC NUCLEAR PORE COMPLEX DURING MITOSIS.
 CC -1- SIMILARITY: THE REPEAT REGION COMPOSED OF PENTAPEPTIDE REPEATS
 CC SEPARATED BY SER/THR-RICH DOMAINS IS SIMILAR TO THAT OF YEAST
 CC NUP1, NSP1, POM 121 AND MAMMALIAN P62.
 CC -1- SIMILARITY: CONTAINS 4 RANBP2-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; L06821; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR001876; znf-RanBP.
DR Pfam; PF00641; znf-RanBP; 4.
DR SMART; SM00547; znf_RanBP; 4.
DR PROSITE; PS01358; ZF_RANBP2_1; 4.
DR PROSITE; PS50199; ZF_RANBP2_2; 4.
DR Nuclear protein; Transport; Repeat; Zinc-finger; DNA-binding;
KW Phosphorylation.
FT DOMAIN 4 15 GLY-RICH.
FT 442 446 POLY-GLY.
FT ZN_FING 657 687 RANBP2-TYPE 1.
FT ZN_FING 721 750 RANBP2-TYPE 2.
FT ZN_FING 789 818 RANBP2-TYPE 3.
FT ZN_FING 846 875 RANBP2-TYPE 4.
SQ SEQUENCE 1468 AA; 152823 MW; C3DFD9697C556A7C CRC64;

Query Match 6.38; Score 82.5; DB 1; Length 1468;
Best Local Similarity 23.38; Pred. No. 45;
Matches 54; Conservative 36; Mismatches 103; Indels 39; Gaps 8;
QY 32 RHHYEKSTPALFSP--KLTRTSLSQRLGNGSAN-----ESWTEVSWPCRNQSSEKAI 84
DB 275 RQNKVRSTP-YQAPVRRQMKAKQLNAQSYGVTSIARRILQSLKMSPLADAKRIPSAV 333
QY 85 N-----LGKKSSWKAFFGVVEKEDSQ-----STPAKVSAGQRTLEYODSHSOOWS 131
DB 334 SSPINSPLDKRGIDSTYQAKKEKVDSDQPPYQRLMTPKPKVSIATNRTYVFKPSLTPSGD 393
QY 132 RCLSNVEQCLEHEAVDPKVISIANRVAELIVSWPPQATQAGGFKSEIFVTEGLSPDLQ 191
DB 394 LRKTNORIDKKNSTVDEKNISRONREQESGFSYPNFSIPANGLSSG---VGGGGCKMR 450
QY 192 GHVPVASSSKKDEFEQLAKTYELKYSGDQLERKDTAFIPPLVDTSIOGF 243
DB 451 ERTTHVAVSKPSEEEV---EVPLLPQ-----ISLPISSSSLPTF 487

Search completed: June 19, 2002, 16:40:08
Job time: 381 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: June 19, 2002, 16:33:40 ; Search time 31.42 Seconds
(without alignments)
770.672 Million cell updates/sec

Title: US-09-771-961-4
Perfect score: 1316
Sequence: 1 MCGTSCDLEIFPLDDDLN.....IPLVDTISIGFPPDGLMACI 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	99.5	7.6	529	2 B23595	DNA nucleotidylexo
2	97	7.4	450	2 T37702	hypothetical prote
3	95	7.2	1083	2 T00790	ubiquitin-specific
4	90	6.8	433	2 S63143	cell division cont
5	89.5	6.8	647	2 G72346	hypothetical prote
6	88.5	6.7	1131	2 A49393	activator 1 large
7	88	6.7	486	1 A35667	Ty transcription a
8	87.5	6.6	287	2 T40340	hypothetical prote
9	87	6.6	347	1 A23595	DNA nucleotidylexo
10	87	6.6	1660	2 A84647	hypothetical prote
11	85.5	6.5	464	2 S75831	hypothetical prote
12	85.5	6.5	889	2 T47311	hypothetical prote
13	85	6.5	1080	2 T00587	probable ubiquitin
14	85	6.5	1468	2 S58250	DNA-directed DNA p
15	84.5	6.4	2875	1 KRVTUW	genome polyprotein
16	84	6.4	336	2 T14437	inner envelope mem
17	83.5	6.3	519	2 AD0663	hypothetical prote
18	83.5	6.3	1284	2 T13168	hypothetical prote
19	83	6.3	332	2 H90576	probable potassium
20	83	6.3	1190	2 P86677	lipoprotein d [imp
21	83	6.3	1193	2 P83264	pyruvate-flavodoxi
22	83	6.3	1265	2 S17675	valine--trna ligas
23	83	6.3	1447	2 T00530	hypothetical prote
24	82.5	6.3	869	2 G82499	clpB protein VCA01
25	82.5	6.3	1148	2 A49651	replication factor
26	82.5	6.3	1468	2 A44345	nucleoporin - rat
27	82.5	6.3	2241	2 T20971	hypothetical prote
28	82.5	6.3	2261	2 T20978	hypothetical prote
29	82	6.2	1121	2 I38127	phosphoprotein pho

30	82	6.2	1123	2 A72311	conserved hypothet
31	81.5	6.2	372	2 G96818	FKK20.3 [imported]
32	81.5	6.2	583	2 T16007	hypothetical prote
33	81.5	6.2	663	2 T49685	RNA polymerase II
34	81	6.2	715	2 D85087	hypothetical prote
35	81	6.2	890	2 G64740	[protein-Pil] urid
36	81	6.2	1647	2 T49412	hypothetical prote
37	80.5	6.1	333	2 T23324	hypothetical prote
38	80.5	6.1	851	2 AG2469	hypothetical prote
39	80.5	6.1	1653	2 A36349	clathrin heavy cha
40	80.5	6.1	1722	1 I78879	retinoblastoma bin
41	80	6.1	178	2 S29886	ribosomal protein
42	80	6.1	427	2 T26103	hypothetical prote
43	80	6.1	496	2 G64527	hypothetical prote
44	80	6.1	690	2 E84945	glycine--trna lig
45	80	6.1	1075	2 B96508	hypothetical prote

ALIGNMENTS

RESULT 1
B23595
DNA nucleotidylexotransferase (EC 2.7.7.31) - mouse
N:Alternate names: terminal addition enzyme; terminal deoxyribonucleotidyltransferase
C:Species: Mus musculus (house mouse)
C:Date: 02-Jun-1988 #sequence_revision 02-Jun-1988 #text_change 18-Jun-1999
C:Accession: B23595; S30235
R:Koizumi, O.; Yokota, T.; Kageyama, T.; Hirose, T.; Yoshida, S.; Arai, K.
A:Title: Isolation and characterization of bovine and mouse terminal deoxynucleotidyl
Nucleic Acids Res. 14, 5777-5792, 1986
A:Reference number: A93633; MUID:8626588
A:Accession: B23595
A:Molecule type: mRNA
A:Residues: 1-529 <KOI>
A:Cross-references: GB:X04123; NID:954765; PIDN:CAA27735.1; PID:954766
R:Doyen, N.; d'Andon, M.F.; Bentolila, L.A.; Nguyen, O.T.; Rougeon, F.
A:Title: Isolation and characterization of bovine and mouse terminal deoxynucleotidyl
Nucleic Acids Res. 21, 1187-1191, 1993
A:Reference number: S30235; MUID:93219079
A:Accession: S30235
A>Status: translation not shown
A:Residues: 1-25, 'M', 27-98, 'F', 100-192, 'R', 194-286, 'O', 288-308, 'E', 310-366, 'D', 368-44
A:Molecule type: mRNA
A:Residues: 1-25, 'M', 27-98, 'F', 100-192, 'R', 194-286, 'O', 288-308, 'E', 310-366, 'D', 368-44
C:Superfamily: DNA nucleotidylexotransferase
C:Keywords: alternative splicing; magnesium; nucleotidyltransferase; nucleus

Query Match
Best Local Similarity 7.6%; Score 99.5; DB 2; Length 529;
Matches 54; Conservative 40; Mismatches 93; Indels 43; Gaps 11;

QY 8 DLEIPLDDDLNTIEFKILAYTRHHVFKSTPALFSPKILRTSLSGRGJGNCANSBW 67
DB 170 DALDLIAENDELRENGSCIAFMGASSVLKSLPPI-----TSMKIDTEGI-PCLDKVK 222

QY 68 TEVSWPCRNSSQSE-KAINIGKKSSKAF---FGYEKEDSOSTPAKVSAGOGTIL-EY 122
DB 223 STIEGIEEGESEKAVLNDERKSKFLTSYGVGLK-----TAEMFRMGFTLSKI 277

QY 123 QSHSQQWSSRC-----LSNVEQCLEHEAVDPKVISIANRVAETIVYSW-PPQATQGG 174
DB 278 QSDKSLRFTKMKAGFLTYEDLVSCVR---PPAQVSMIVKRAVVTLPDALVTMGG 333

QY 175 FSKSEIVTEGSLFQLOGH---VPVASSSKKDEEOLIAKIVELTKYSG 220
DB 334 FRKG-----KMTGHDVDFLTSPKATEDEEQLLHKVTHFWKQGG 373

RESULT 2
T37702
hypothetical protein SPAC15A10.02 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T37702
R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1997
A:Reference number: Z21738
A:Accession: T37702
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-450 <MOR>
A:Cross-references: EMBL:Z97208; PIDN:CAB10099.1; GSPDB:GN00066; SPDB:SPAC15A10.02
A:Experimental source: strain 972h-; cosmid c15A10
C:Genetics:
A:Gene: SPDB:SPAC15A10.02
A:Map position: 1

Query Match	7.4%	Score 97;	DB 2;	Length 450;
Best Local Similarity	28.1%	Pred. No. 1.1;		
Matches	57;	Conservative	26;	Mismatches 94;
			Indels	26;
			Gaps	9

51 RSLSQRLGN---CSANESWTEVSWPCNSQSSSEKAI----NLGKKKSSWKAFEGVYEKE 103
 || : | | | : | : | : : : :
 184 RSLQQLQ--LQNFQASSANGNNTGTSTPVNASTDTRFKSTASTPQLQQLQQAQANAPQRLINPE 244

QY	104	DSQ	---STFAKSAO	---CGRTLEYODSHQOQSRCLSNOCLEHAAVDPKVSIANRAE	159
Db	242	TSSVEFPVGSAAVNSNE	STELATSAIQ	---SGLANVNEKSGQTSPYSAMHLPRVDEKSP	300
QY	160	IYVSMPPOAOAGFESKET	---FTEGEGS---	---POLQGHVPAVASSKK---	207
Db	301	IPFSVPKATLTGGYAGSG	ISGLSTPGLSAPRPHYELDNGNANL	SKKRLHDLILQOIDBEK	366

```
QY      208 ILAKIVELKYSGDQLERKDTAF 230
          |::|||:|
Db      361 IEPVEELLLEIADFEVESVTNF 383
```

RESULT 3
T00790
ubiquitin-specific protease homolog F24L7.8 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00790; D84737
R:Rounsley, S.D.; Lin, X.; Kechum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaud
submitted to the EMBL Data Library, February 1998
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A:Reference number: Z14204

status: translated from GB/EMBL/DBJ
molecule type: DNA
residues: 1-1083 <R0U>
A: Cross-references: EMBL:AC003974; NID:q2914668; PTD:q2914655

A; Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.
; Reference number: AB4420; MID:20083487

A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-1083 <STO>
A;Cross-references: GB:AE002093; NID:g2914695; PIDN:AAC04485.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g32780; F24L7.8
A;Map position: 2
A;Introns: 416/2; 1004/2

Query Match	7.28;	Score 95;	DB 2;	Length 1083
Best Local Similarity	21.88;	Pred. No. 5.3;		

	Matches	62,	Conservative	39,	Mismatches	94,	Indels	90,	Gaps	13,
Qy	8	DLEELPLDDDD	DLNTEERKILAYTTRHHVKEKTPAL	-----	FSPKLLRTSSLSQRGIG	59				
Db	766	DSDPVLVDSBSPVSVDRKLAQFTTHETLSEDMAMHCNCSKNLKLQRLREKRRTKEGGS	845							
Qy	60	NCSANESWTEVSW-PCRNSSOSEKAINLGKKSSWKA	-----	FGVVE	101					
Db	846	NKMWVENGASASFDECRDPSLQNCIDL	---ENGKKAAPPTIKLPNCKKEESAIDDFVG	902						
Qy	102	KEDSGSTP----	AKXSAGQRTLEVO----	DSHSQMSKRLSVBECLEHADVAVISIA	154					
Db	903	EENIQAPITSTYETPLLGGETISSQPSADNCEEMW	-----	EDLAVDSBEV-IV	950					
Qy	155	NKVA-EIVYSMPPOAT	-----	QAGGFKSEIEFTVGLSFOLOGHPV	-----	196				
Db	951	KRDARKKYLINAPVLVLIHLKRFSGODARGVSK	-----	LSGVHPDQEFIDLSK	999					
Qy	197	---ASSSKKDEEQLAKIVELL	-----	KYSGDLERKDT	228					
Db	1000	YMDTCSEDEFPVYRLAGVHEHGLGMSRGHYSYIRGGHKEKRD	1044							

RESULT 4
S63143
cell division control protein KAR1 - Yeast (Saccharomyces cerevisiae),
Nucleotide sequence analysis: 1611 nucleotides, 540 aa.

N;Alternate names: protein N1611; protein YNL188w

C;Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 29-Oct-1999

R;Obermaier, B.: Piravandi, E.: Rinke, M.: Domdey, H.
C/accession: 505145, 420455

submitted to the Protein Sequence database, April 1996

A;Accession: S63143

A;Residues: 1-433 <OBE:

A; Experimental source: strain S288C

R;rose, M.D.; Fink, G.R.
Cell 48, 1047-1060, 1987

A;Title: KARL, a gene required for function

A;Accession: A26455

A/Residues: 1-198, 'V', 200-433 <ROS>

A; Cross-References: GB:M15683; NID:g1/17/6; PL

A;Gene: SGD:KAR1

A;Map position: 14L

F;413-429/Domain: transmembrane #status predicted <TMM>

Query Match	6.88;	Score 90;	DB 2;	Length 433;
-------------	-------	-----------	-------	-------------

Matches 55; Conservative 43; Mismatches 107; Indels 52; Gaps 10;

QY 8 DLEIPLDDDL-NTIEFKILAYTRHHVFKSTPALFSPKLLRTRSLSQRL---GNCSA 63

Db 131 DEBNYKIDEDNTENEI.OE-----TPKTKFASTI.RS.SI.GORNYI.NTBPKS 176

1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 26

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	52
--	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	----

```

0Y      123  ---QDSHSQOWSRCLSN---VEQCL-----EHEAVDPKVISIAN-----RVAELV 161
      . . . . .

```

Db 237 LQKKRELIESKWHRLLFHDKKMEKKLESRLREYERKRMPRGTDVSSSEQDNSFKISTPT 296

QY 162 YSW-----PPQATQAGF-----KSEIFVTEGLSFOLQGHVPVASSSKKDEEEQIL 209

Db 297 KSYVSLQKPLPNI¹SAMNNFNDVTDNKEKEETNN²NI³LKFOAORDP⁴LOILOSEIEMHT⁵KL 356

submitted to the Protein Sequence Database, August 1994

A:Reference number: S45893

A:Accession: S45950

A:Molecule type: DNA

A:Residues: 1-486 <AND>

A:Cross-references: EMBL:Z35952; NID:9536345; PIDN:CAA85028.1; PID:9536346; GSPDB:GN0000

C:Gene: SCD:TEC1; MIPS:YBR083w

A:Cross-references: SGD:S0000287; MIPS:YBR083w

A:Map position: 2R

C:Superfamily: Ty transcription activator TEC1; TEA DNA-binding domain homology

C:Keywords: DNA binding; nucleus; transcription regulation

E:123-193/Domain: TEA DNA-binding domain homology <TEA>

Query Match 6.7%; Score 88; DB 1; Length 486;

Best Local Similarity 22.6%; Pred. No. 7.6;

Matches 58; Conservative 36; Mismatches 95; Indels 68; Gaps 12;

3 STSGCDLEETPLDDDDLTIEFKILAYTRHHVFK-----STPALFSRK 46

48 STNACVKNKA---EDNNLID-----THPQFELVMTGLGAKSDCLKSPSKAFTTDK 96

47 LKTR-----SLSGRLGNCANESWTEVSPCRNSQSEK-----AINLKKKSKMK 94

97 QRKNVFNISVSNVFPQSSSTSTTE-SMTIGCDKMKSEKAEALRLIMKNGTTKI 155

95 ----AFGEVKEKEDOSTPAKVASOQORTLEYQDSHOSQMSCLSN-VQCLEHEAVDPK 149

156 KINNANFG---RNEILSLYIKHKTNERTTKQIISHIOVWKTTONKIKDLSLSSKEKE 212

150 VISIANRVA-----EIVYSWPPQATQAGFKSEKLFV---TEGLSFQQL 191

213 LHLIEHGAQTENSMLFQDIFEIITDSL-PSVSDSGSLPKMLIYVNSNSGLSVHAK 270

192 GAVPVASSSKKDEEQT 208

271 LKLPITASNKKIENFI 287

RESULT 8

T40340

Hypothetical protein SPBC3B9.01 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T40340

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z21922

A:Accession: T40340

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-287 <MO>

A:Cross-references: EMBL:AL022070; PIDN:CAAL17781.1; GSPDB:GN00067; SPDB:SPBC3B9.01

C:Experimental source: strain 972h-; cosmid C3B9

C:Genetics:

A:Gene: SPDB:SPBC3B9.01

A:Map position: 2

Query Match 6.8%; Score 87.5; DB 2; Length 287;

Best Local Similarity 21.2%; Pred. No. 4.1;

Matches 52; Conservative 41; Mismatches 113; Indels 39; Gaps 10;

12 IPDDDDLTNIEFKILAYTRH--HFKSTPALSPKLLRTSLSSORGLGNCANSANSWTE 69

56 VPLDQKE---IAPDNLEMLVEHIDNANNLVPLQLMPRLIKOLESPSESTLRLAA--WT- 108

70 VSMPCNNSOSSEKAINLGKSSWKAFFGVVEKEDOSTPAKV-----SAOGORTLEY 122

109 INTAVGNPNRSGQAL---IENDGKILIEGALKEDSDTKNNVLAITSKLINENGIAL 165

123 QDSHOSQMSRCLSNVEQCLEHEAVDPKVIIS--ANRVAEIVYSWPPQATQAGFKSEK 180

Db 166 LDKIPNSWEMLEIIE--LKHSVMTKRVIFEFYALLIQE-----DKSKOI 208

Qy 181 FVTBGLSFOLGHPVAVASSKKDEEQIILAKIYELLYKSDOLEKRTAFPIPLVPTSI 240

Db 209 ILQKHEHPEKRYOFSLSHSVDEDCVTKSLHTLTLFQKNKVSANTNELSLVQPKS 268

Qy 241 QGFPQ 245

Db 269 E-FPE 272

RESULT 9

A23595

DNA nucleotidyltransferase (EC 2.7.7.31) long form - bovine

N:Alternate names: terminal addition enzyme; terminal deoxyribonucleotidyltransferase

C:Species: Bos primigenius taurus (cattle)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A23595; A60467; A05052; A29948; A32198; PC2127; PC2128; I45884

R:Koiwai, O.; Yokota, T.; Kageyama, T.; Hirose, T.; Yoshida, S.; Arai, K.

Nucleic Acids Res. 14, 5777-5792, 1986

A:Title: Isolation and characterization of bovine and mouse terminal deoxynucleotidyl

A:Reference number: A93633; MUID:86286588

A:Accession: A23595

A:Molecule type: mRNA

A:Residues: 1-302,321-510,520-547 <KOI>

A:Cross-references: GB:X04122; NID:9767; PIDN:CAA27734.1; PID:9768

A>Note: It is uncertain whether Met-1 or Met-12 is the initiator

R:Evans, R.K.; Besch, C.M.; Coleman, M.S.

Biochemistry 28, 713-720, 1989

A:Title: Photoaffinity labeling of terminal deoxynucleotidyl transferase. Identificat

A:Reference number: A60467; MUID:89229071

A:Accession: A60467

A:Molecule type: protein

A:Residues: 143-179,'X',181-187,'XXXXXX',194-199,'XX',202-230,'XXX',234-303,322-363,'

R:Peterson, R.C.; Cheung, L.C.; Mattaliano, R.J.; Chang, L.M.S.; Bollum, F.J.

Proc. Natl. Acad. Sci. U.S.A. 81, 4363-4367, 1984

A:Title: Molecular cloning of human terminal deoxynucleotidyltransferase.

A:Reference number: A93995; MUID:84272638

A:Accession: A05052

A:Molecule type: protein

A:Residues: 335-348,'X',350-355,366-377,'XX',380-381,'X',383-386,521-526,'X',528-533,

R:Pandey, V.; Modak, M.J.

J. Biol. Chem. 263, 3744-3751, 1988

A:Title: Biochemistry of terminal deoxynucleotidyltransferase. Affinity labeling and

A:Reference number: A29948; MUID:88153669

A:Accession: A29948

A:Molecule type: protein

A:Residues: 221-226,'X',228-231,'X',235-249 <PAN>

R:Pandey, V.N.; Modak, M.J.

J. Biol. Chem. 264, 867-871, 1989

A:Title: Biochemistry of terminal deoxynucleotidyl transferase. Identification and unl

A:Reference number: A32198; MUID:89093157

A:Accession: A32198

A:Molecule type: protein

A:Residues: 221-226,'X',228-231,'X',235-249 <PA2>

R:Takahara, K.; Hayashi, N.; Fujita-Sagawa, K.; Morishita, T.; Hashimoto, Y.; Noda, A

Biosci. Biotechnol. Biochem. 58, 786-787, 1994

A:Title: Alternative splicing of bovine terminal deoxynucleotidyl transferase cDNA.

A:Reference number: PC2127; MUID:94264411

A:Accession: PC2127

A:Molecule type: DNA

A:Residues: 303-320 <TA2>

A:Cross-references: GB:D17451

A:Accession: PC2128

A:Molecule type: DNA

A:Residues: 511-519 <TA2>

A:Cross-references: GB:D17452

R:Koiwai, O.; Kaneda, T.; Morishita, R.

Biochem. Biophys. Res. Commun. 144, 185-190, 1987

A:Title: Analysis of human terminal deoxynucleotidyl transferase cDNA expressible in

A:Reference number: I45884; MUID:87213162

A:Accession: I45884

A:Map position: 14L
C:Superfamily: herpesvirus DNA-directed DNA polymerase
C:Keywords: nucleotidyltransferase

Query Match 6.5%; Score 85; DB 2; Length 1468;

Best Local Similarity 25.9%; Pred. No. 62;
Matches 51; Conservative 29; Mismatches 93; Indels 24; Gaps 8;

QY 65 ESMTEVSWPCRNSSQSEKAINLGKKSSWKAFFGVKEKEDSOSTPAKVSAGQRTLEYODSHSOQWSRCLIS 124
DB 74 EDWREVD---NSSDEPDGNTLASKDSKRKK--NIKREKHQITDMLRTQHSKSTLLAHA 127
QY 125 SHSOQWSRCLISVEOCL---EHEAVDPKVISIANRVAEIVYSWPPQATQAGFKS--KE 179
DB 128 KKSOKKSIPIDNFDILGEFESGEVEKPNILPSKIRENLNSPTSE-----FKSSIKR 181
QY 180 IFVTGLSPQ--LQGHVPVASSSKKDEBQILAKIVL---LKYSQDLERKDTAFIPI 233
182 VNGNDESHDAGISKKVKIDPDSDTKYLESSPLKLSRKLYANDVDLDDV-ENS 240
QY 234 PLVDTSIGFPQDGLMA 250
DB 241 PVATKRONVLDITLLA 257

RESULT 15

RRVUTW
genome polyprotein - tomato spotted wilt virus (strain BR-01)
N:Alternate names: L protein
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: tomato spotted wilt virus
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 29-May-1998
C:Accession: J01335
J:de Haan, P.; Kormelink, R.; de Oliveira Resende, R.; van Poelwijk, F.; Peters, D.; Gol
J. Gen. Virol. 71, 2207-2216, 1991
A:Title: Tomato spotted wilt virus L RNA encodes a putative RNA polymerase.
A:Reference number: J01335
A:Accession: J01335
A:Molecule type: genomic RNA
A:Residues: 1-2875 <DEH>
C:Genetics:
A:Map position: segment L
C:Superfamily: bunyavirus RNA-directed RNA polymerase
C:Keywords: nucleotidyltransferase

Query Match 6.4%; Score 84.5; DB 1; Length 2875;

Best Local Similarity 22.0%; Pred. No. 17e+02;
Matches 40; Conservative 40; Mismatches 73; Indels 29; Gaps 7;

QY 80 SEKAINLGKKSSWKAFF---GVVEKEDSOSTPAKVSAGQRTLEYODSHSOQWSRCLIS 135
DB 1760 NENALN--KMSYKDFSKLYPNLKKNEDLYKSTNKLKIDEDAVLE---EDELKTKIAS 1812
QY 136 NVEQCLEHEAN--DPKVISIAN-----RVAEIVYSWPPQATQAGFKSKEIFVTEGLS 187
DB 1813 SLEMESYHDIMIKNPETLLAPLNDRLFLSQLFMYTSPSKRWOLNSOSTEKLALDRVLR 1872
QY 188 FOLQGHVPVASSSKKDEBQILAKIVELIKYSGD-----QLEKRDPAF--IPIPLVD 237
DB 1873 SKARTEFVNISSTVAMTEENNEKKILEMKLPDLDSYCSFKTCVNLVIKIDVNFSLIPILD 1932
QY 238 TS 239
DB 1933 SA 1934

Search completed: June 19, 2002, 16:33:44
Job time: 77 sec

